DOCUMENT RESUME

ED 197 416 CS 503 255

TITLE	Status Report on Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, July 1-December 31, 1980.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Inst. of Child Health and Human Development (NIH), Bethesda, Md.: National Inst. of Education (DHEW), Washington, D.C.: National Science Foundation, Washington, D.C.
REPORT NO	SR-63/64
PUB DATE	80
GRANT	HD-01994: NIE-G-0178: NSF-MCS79-16177
NOTE	347p.
EDRS PRICE	MF01/PC14 Plus Postage.
DESCRIPTORS	*Acoustic Phonetics: *Articulation (Speech): Auditory Perception: *Communication Research: Consonants: *Oral Language: Second Languages: Speech Handicaps: *Speech Skills: Spelling: Structural Analysis (Linguistics)

ABSTRACT

Research reports on the nature of speech, instrumentation for the investigation of speech, and practical applications of research are included in this status report for the July 1-December 31, 1980, period. The 19 reports deal with the following topics: (1) the structuring of langauge: (2) whether movements are prepared in parts: (3) velopharyngeal function: (4) speech perception without traditional speech cues: (5) the influence of preceding liquid on stop consonant perception: (6) perceptual assessment of fricative-stop coarticulation: (7) two strategies in fricative discrimination: (8) context sensitivity and phonetic mediation in categorical perception: (9) bidirectional contrast effects in the perception of vc-cv sequences: (10) perception and production of two stop-consonant sequences: (11) the naming of words in Kana and in Kanji: (12) the Roman and Cyrillic alphabets of Serbo-Croatian: (13) lexical decision in a phonologically shallow orthography: (14) representation of inflected nouns in the internal lexicon: (15) a word superiority effect in a phonetically precise orthography: (16) laryngeal activity in Icelandic obstruent production: (17) laryngeal adjustments in Japanese voiceless sound production: (18) articulatory control in a deaf speaker: and (19) acoustic factors contributing to categorical perception. (FL)

****

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

****

ED197416

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

SR-63/64 (1980)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 July - 31 December 1980

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

CS503255

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994

National Institute of Child Health and Human Development
Contract N01-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant MCS79-16177
Grant PRF8006144

National Institute of Neurological and Communicative Disorders and Stroke
Grant NS13870
Grant NS13617

National Institute of Arthritis, Metabolism, and Digestive Diseases
Grant AM25814

National Institute of Education
Grant G-80-0178

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper,* Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey, Treasurer
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Peter Alfonso*
Thomas Baer
Alan Bell[3]
Fredericka Bell-Berti*
Catherine Best+
Gloria J. Borden*
Susan Brady*
Robert Crowder*
William Ewan*
Carol A. Fowler*
Louis Goldstein*
Katherine S. Harris*
Alice Healy*
Kiyoshi Honda[1]
Daisy Hung
Leonard Katz*
Scott Kelso
Andrea G. Levitt*
Isabelle Y. Liberman*
Leigh Lisker*
Anders Löfqvist[2]
Virginia Mann+
Charles Marshall
Ignatius G. Mattingly*
Nancy S. McGarr*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Donald P. Shankweiler*
Michael Studdert-Kennedy*
Michael T. Turvey*
Ovid Tzeng[4]
Mario Vayra[5]
Robert Verbrugge*

Technical and Support Staff

Eric L. Andreasson
Elizabeth P. Clark
Vincent Gulisano
Donald Hailey
Terry Halwes
Sabina D. Koroluk
Agnes M. McKeon
Nancy O'Brien
Marilyn K. Parnell
William P. Scully
Richard S. Sharkany
Leonard Szubowicz
Edward R. Wiley
David Zeichner

Students*

Claudia Carello
Tova Clayman
David Dechovitz
Steven Eady
Jo Estill
Laurie Feldman
Hollis Fitch
Carole E. Gelfer
David Goodman
Janette Henderson
Charles Hoequist
Kenneth Holt
Robert Katz
Aleksandar Kostić
Peter Kugler
Anthony Levas
Harriet Magen
Suzi Pollack
Patti Jo Price
Sandra Prindle
Brad Rakerd
Daniel Recassens
Arnold Shapiro
Susan Smith
Louis G. Tassinary
Janet Titchener
Emily Tobey-Cullen
Betty Tuller
Douglas Whalen
Deborah Wilkenfeld

---

*Part-time
[1] Visiting from University of Tokyo, Japan
[2] Visiting from Lund University, Sweden
[3] Visiting from University of Colorado
[4] Visiting from University of California, Riverside
[5] Visiting from Scuola Normale Superiore, Pisa, Italy
+NIH Research Fellow

SR-63/64 (1980)
(July-December)

CONTENTS

I. Manuscripts and Extended Reports

The structuring of language: Clues from the differences between signed and spoken language--Michael Studdert-Kennedy and Harlan Lane 1

Are movements prepared in parts? Not under compatible [natural] conditions--David Goodman and J. A. Scott Kelso 9

Velopharyngeal function: A spatial-temporal model--Fredericka Bell-Berti 41

Speech perception without traditional speech cues--Robert E. Remez, Philip E. Rubin, David B. Pisoni, and Thomas D. Carrell 67

Influence of preceding liquid on stop consonant perception--Virginia A. Mann 79

Perceptual assessment of fricative-stop coarticulation--Bruno H. Repp and Virginia A. Mann 93

Influence of vocalic context on perception of the [ʃ]-[s] distinction: IV. Two strategies in fricative discrimination--Bruno H. Repp 121

Context sensitivity and phonetic mediation in categorical perception: A comparison of four stimulus continua--Alice F. Healy and Bruno H. Repp 139

Bidirectional contrast effects in the perception of VC-CV sequences--Bruno H. Repp 157

Perception and production of two-stop-consonant sequences--Bruno H. Repp 177

Words written in Kana are named faster than the same words written in Kanji--Laurie B. Feldman and M. T. Turvey 195

Some experiments on the Roman and Cyrillic alphabets of Serbo-Croatian--G. Lukatela and M. T. Turvey 203

Lexical decision in a phonologically shallow orthography--G. Lukatela, D. Popadić, P. Ognjenović and M. T. Turvey 227

Representation of inflected nouns in the internal lexicon--G. Lukatela, B. Gligorijević, A. Kostić and M. T. Turvey 243

A word superiority effect in a phonetically precise orthography--G. Lukatela, B. Lorenc, P. Ogjenović and M. T. Turvey 263

Laryngeal activity in Icelandic obstruent production--Anders Löfqvist and Hirohide Yoshioka 275

Laryngeal adjustments in Japanese voiceless sound production--Hirohide Yoshioka, Anders Löfqvist and Hajime Hirose 293

Articulatory control in a deaf speaker--Nancy S. McGarr and Katherine S. Harris 309

Acoustic factors that may contribute to categorical perception--Janet G. May 333

II. Publications 351

III. Appendix: DTIC and ERIC numbers (SR-21/22 - SR-59/60) 353

I. MANUSCRIPTS AND EXTENDED REPORTS

THE STRUCTURING OF LANGUAGE: CLUES FROM THE DIFFERENCES BETWEEN SIGNED AND SPOKEN LANGUAGE*

Michael Studdert-Kennedy+ and Harlan Lane++

Abstract. The formational structures of signed and spoken language are compared in terms of both their phonemes, or primes, and their features. The comparison leads to the suggestion, first, that the two levels of sublexical structure in both languages provide a kind of impedance match between an open-ended set of meaningful symbols and a decidedly limited set of signaling devices; and, second, that while speech draws on a degree of parallel organization to implement a sequential linguistic structure, sign implements a parallel linguistic structure by a partially sequential organization of its gestures. The differences seem to arise because the hands have more degrees of motor freedom than the mouth and/or because the spatial patterns available to sight afford a richer simultaneous structure than the temporal patterns available to hearing.

INTRODUCTION

If we assume that the two modes of communication, speaking and signing, draw on shared cognitive structures, then systematic differences between spoken and signed languages must result from differences in modality, while similarities may reflect either cognitive properties of language or cross-modality invariances in its implementation. It is such invariances--of motor organization, of perception, or of representation in memory--that may constrain the structure of language.

A fundamental discovery of recent years, due to systematic analysis of American Sign Language (ASL) (Stokoe, Casterline, & Croneberg, 1965; Klima & Bellugi, 1979) is that a dual pattern of syntax and form characterizes signed no less than spoken language. Although a two-leveled structure is often said to be distinctive of human language, its origin and function are seldom discussed. The functional advantages of the one level, syntax, with its

---

*To be published in Bellugi, U. and Studdert-Kennedy, M. (Eds.), Signed Language and Spoken Language: Biological Constraints on Linguistic Form. Dahlem Konferenzen. Weinheim/Deerfield Beach, Fl./Basel: Verlag Chemie, 1980.

+Also Queens College and Graduate Center, City University of New York.

++Northeastern University, Boston, Massachusetts.

Acknowledgment: We gratefully acknowledge J. Kegl's important contributions to this paper. Preparation of the paper was supported, in part, by NICHD Grant No. HD 01994 to Haskins Laboratories, and by a grant from the Division of Behavioral and Neural Sciences, National Science Foundation, to Northeastern University.

powers of unambiguous predication, repeated recursion, and so on, are apparent. As for its origin, it is not inconceivable that syntactic structure evolved by exploiting neural networks already developed for hierarchical control of motor behavior, but this is a matter well beyond the scope of present speculation. The function of the second level, formational structure, is less obvious and its origin may be more amenable to investigation.

Consider a language with a syntax, that is, rules for forming utterances by combining meaningful elements, but with no phonology, no rules for forming meaningful elements from smaller units. Meaningful elements would then be holistically distinct signals, devoid of systematic interrelations. If the lexicon were iconic, its limits would be set by the human capacity to represent--obviously a more severe constraint for acoustic than for visual form--and abstraction would be difficult, if not impossible. If the elements were not iconic but arbitrary, the lexicon would again be limited, because the number of holistically distinct signals that humans can form at a reasonable rate, vocally or manually, and perceive by ear or by eye, is small. (Most vertebrate communication systems dispose of fewer than 40 distinct signals.) Of course, the lexicon could be enlarged by reduplication of elements (the first step toward structure, incidentally), but this would be a cumbersome solution, making, in the end, prohibitive demands on memory. While a modest lexicon does not preclude a productive syntax, and while listeners will submit to a surprising degree of homonymity (Klima, 1975), it is clear that a lexicon adequate to human cognitive demand could not be constructed without recourse to submorphemic structure.

What are the requirements of such structure? Perceptually, they are simple. First, signals must be attuned to psychophysical capacity. Thus, speech sounds are concentrated in the center of the audiogram and visual information during signed communication tends to concentrate around the observer's line of sight--larger signs with more ample movement occur in the periphery of the visual field, while those requiring finer discrimination occur closer to the fovea. Also, boundaries among phoneme or prime categories must be placed at points of adequate discriminability. There is some evidence for the psychophysical determination of at least some such boundaries in speech, although not yet in sign. But the strongest perceptual demand is that the submorphemic units be so compacted that they place minimal demands on short-term storage before lexical access transfers the processing load to syntactic and semantic mechanisms.

From this perceptual demand spring the motor requirements. The signaler must have at his command a rapid and precise peripheral mechanism with enough degrees of freedom for a fair repertoire of distinct gestures. Speed and precision call for a flexible system with a high degree of central neural coordination. Presumably, it is no accident that cerebral localizations of manual control and linguistic function are associated. Manual and vocal systems probably draw on common principles and mechanisms of motor control.

SERIES-PARALLEL DIFFERENCES IN MORPHEME STRUCTURE

From a linguistic perspective, there are obvious differences between the structures of speech and sign. Most salient are the different ways in which they combine their meaningless units (phonemes, primes). Why does speech combine its units in series, sign in parallel? Or, why does ASL not prefer to fingerspell (using arbitrary units unrelated to those of speech), and why does speech not prefer to stack its units in simultaneous bundles?

The most obvious response is to attribute the series-parallel difference to perception and to the differences between sound and light. The distinction is not clear-cut, but sound does entail primarily a temporal, light primarily a spatial distribution of energy. The distinctive gestures of speech and sign seem to be adapted to the medium through which they are conveyed. For example, the spatial distinctions of tongue height among the whispered high-front-vowel /i/, the fricative /s/, and the stop /t/ (as in east) are a matter of a few millimeters, barely perceptible when viewed spatially by X-ray, but highly discriminable when transduced into the temporal array of sound. Similarly, the extensive use of space in sign language reflects the adaptation of the language to the visual medium. Yet, the visual system is clearly comfortable with a sequential display (ASL compounding, infixing, indexing; negative, topic, and aspect marking) and the auditory system readily discriminates among simultaneous properties (tones, nasalization, stress).

Motor as well as perceptual constraints may underlie the series-parallel difference between modalities. Note first that speech is not entirely sequential. Each phone is formed from a roughly "simultaneous bundle" of articulatory features, and each feature is reflected in the signal by at least some more-or-less simultaneous, often spectrally dispersed, acoustic cues. (We use the term "feature" loosely to refer to an isolable property of a gesture, such as tongue root advancement, glottal closure, or velar opening. We are not here concerned with the abstract features of phonology, each of which may be compounded from several articulatory features. We do, however, propose that, in the last analysis, the feature structure of phonology derives from the feature structure of its modality of expression.)

The feature structure of speech is, in large degree, a consequence of the anatomy and physiology of the vocal tract. The active articulators, carrying the major phonetic load (larynx, tongue, jaw, lips, velum), are few, and each has relatively few discriminable states (here again perception impinges). Moreover, none of the articulators can work in isolation; all are engaged (even if only passively) in the production of any single sound. A sizable repertoire of sound units can therefore only be built by repeated use of the same articulator, and of a particular action of that articulator, in more-or-less simultaneous combination with the several actions of other articulators.

To this extent, speech is no less parallel in form than is sign (see below). We might even wonder why features are not the basic meaningless units of speech and phonemes the basic meaningful elements. Single phonemes are indeed used in many languages to fulfill morphemic functions (interestingly, from the point of view of rate, these are often high-frequency grammatical morphemes). However, if this were general, spoken languages would be reduced to a maximum of roughly a hundred morphemes. This limit is placed because many combinations of features are excluded: they call either for the same articulators or for incompatible actions by different articulators. We cannot specify exactly how many combinations are possible without knowing the degrees of freedom of the vocal tract--knowledge that awaits a fuller understanding of its motor control. However, we can estimate the upper limit from the maximum number of phonemes found in any single language, and this is roughly a hundred. Thus, limits on the vocal apparatus force speech, first into a featural structure of its units (phonemes), then into concatenation of those units, in order to achieve an appreciable repertoire of semantic elements.

Yet concatenation carries a penalty: neighboring units are formed by the same small set of articulators, and articulators are limited in the rate at which they can switch from one action to another. Here again, the feature structure of speech permits a solution: carry-over of feature values from one phoneme to the next (Cooper, 1972). The opening gesture that releases a consonant is itself a property of the following vowel, while the vowel is, in turn, a precondition of the following consonantal constriction. Thus, as one phonetic unit is produced, the unengaged or partially engaged components of a later unit are being activated: in the word bought, for example, lips round for medial vowel, before they open to release the initial labial consonant, and tongue tip rises for final alveolar closure, while its root is still backed and lowered for the preceding vowel. Thus, the fundamental element of spoken language, the consonant-vowel syllable, is formed by the intricate, overlapping gestures associated with both simultaneous and sequential articulatory features.

Pursuing the series-parallel difference, let us apply this line of reasoning to sign language. There would be too few signs, as there would be too few words, if each was holistically different from the next. Similarly, the primes (hand configurations, locations, movements) from which signs are constructed draw on a modest number of articulators with relatively few possible states. There would be too few hand shapes if each shape had to be holistically different from every other, too few movements if each movement shared no features with any other, and so on. Thus, we motivate a level of structure below the level of the prime in sign, as in speech.

But now the types of language part ways. The greater degrees of freedom of the signing apparatus and the visual modality allow sign language to transmit its selected combining elements concurrently rather than sequentially. Occasionally, two primes are sequentially adjacent within a sign, like two phonemes in a word. This small set of signs is then subject to severe phonotactic constraints which tend to make the combining elements maximally opposed on major class features. More commonly, sequentially adjacent primes are separated by a morpheme boundary. For both these reasons, we see little sequential coarticulation in ASL. What we find instead is a tendency for simultaneous elements to interact. Movements are reduced, or shifted from arm to wrist, wrist to finger. Handshapes are adjusted to facilitate contact between body parts. For example, the thumb is moved away from its position across the fingers, as in a fist, to permit the knuckles to touch in the two-handed signs MEET and WASH; the index protrudes from the fist, at the second joint, to contact the face at chin or temple in APPLE and ONION, respectively.

However, we should note that these adjustments are not intrinsic to the manual system as the coarticulations of speech are to the vocal apparatus. The unadjusted hand shapes or movements are physically possible, without loss in the rate of information transfer, as the mutual adjustments of consonant constriction and vowel opening are not. In other words, the coarticulation effects of sign language are extrinsic variations, analogous to the presence of aspiration in a syllable-initial English /p/ and its absence in an /sp-/ cluster, rather than intrinsic, as in the spectral and temporal variations that accompany the articulation of a particular consonant before or after different vowels. (For the distinction between extrinsic and intrinsic allophonic variations, see Wang and Fillmore, 1961.)

In short, a comparison of speech and sign leads us to suggest first, that the two levels of sublexical structure in both languages provide a kind of impedance match between an open-ended set of meaningful symbols and a decidedly limited set of signaling devices; and second, that sign transmits the elemental units at both levels in parallel whereas speech transmits phonemes sequentially, features in parallel. This difference seems to arise because the hands have more degrees of motor freedom than the mouth and/or because the spatial patterns available to sight afford a richer simultaneous structure than the temporal patterns available to hearing.

If this account is correct, we may conclude that it is the differences in modality between speech and sign that determine their differences in morpheme structure. Although spoken language may occasionally make lexical distinctions by means of simultaneous variations in, say, spectral structure and fundamental frequency (as in tone languages), for the most part, it is the ordering of elements that specifies the morpheme, so that, whatever coarticulatory interleaving may occur, the basic sequence must be preserved in execution. By contrast, again with some few exceptions, ASL does not use the ordering of elements to distinguish morphemes.

SERIES-PARALLEL DIFFERENCES IN EXECUTION

Yet, as we have already suggested, the series-parallel distinction begins to reverse itself when we examine the detailed processes of execution: parallel processes appear in speech, sequential processes in sign. Thus, Fowler (1979; cf. Öhman, 1966) has argued that coarticulation effects are due not to the spread of features (such as lip-rounding, velar opening, tongue raising) across neighboring segments but to actual simultaneous or coproduction of consonants and vowels. In this view, the neuromuscular synergisms or coordinative structures involved in vowel production are engaged just once at the start of an utterance and then continue to cycle rhythmically with minor adjustments throughout the utterance. On this underlying and relatively slow rhythmic base are superimposed the actions of the distinct and more rapid coordinative structures involved in consonant production. For example, "lip rounding precedes the measured acoustic onset of a rounded vowel, and therefore coarticulates with the preceding consonants... not because the feature [+rounding] has attached itself in the plan to the preceding consonants, but rather because the vowel /u/ is coproduced with them" (Fowler, 1979, p. 61).

This description of articulation as co-occurring coordinative motor structures highlights the resemblance between speech and sign production. The stream of signing can be viewed as the result of coordinative motor structures producing cyclical movements of the arms, on which are superimposed fine movements of the wrists and fingers. The cyclical movements are checked by contact with parts of the body or go unchecked. The dance of the arms on the vertical surface of the body resembles the dance of the tongue on the roof of the mouth. Both systems allow interruption of movement to occur when a moving articulator contacts either a fixed or a movable articulator. If the distance from the waist to the crown is greater than from the lip to the pharynx, the arm is also longer than the tongue, and a long lever is slow to move. If we recall, further, that the proximal stimulus for sign perception is typically some five feet away from the signer, it is not evident that sign has much greater possibilities than speech for simultaneous transmission either motorically or perceptually.

If we suppose then that interfacing speech and sign with their peripheral articulators imposes similar constraints on each, we are led to inquire where in sign language are the temporally organized coproduction effects, such as Fowler's lip rounding example, that we find in speech. If phonemes (feature bundles) are the first level of submorphemic structure and features the second, where are the changes in the feature bundles caused by interleaving one bundle with another, one set of articulatory configurations overlapping another?

Consider the entry in the Stokoe et al. (1965) dictionary for the sign translated in English as LATER. The entry indicates that the phonemically distinct tokens of the three sign aspects that combine simultaneously are (for the dominant hand) L-handshape (as in SHOOT), nodding movement (as in YES), and location on the nonspread flat palm of the nondominant hand (as in CERTIFY). Yet when we look more closely at this example, we are tempted to reorganize the data in such a way that what have traditionally been considered phoneme-like primes are viewed instead as morphemes. Not only are the units involved not meaningless, they are also not fully simultaneous. Rather, they are morphemes that have undergone sequencing and rule-governed alternations. First, the base hand is in the common classifier configuration for flat movable objects (BOOK, PAPER, MIRROR). Let us call it //FMO//. Next, the dominant hand has the pointing configuration used for indexing, for two things pointing at each other (OPPOSE, ARGUE) and for designating units of time (WEEK, MONTH). Call it //POINT//. Finally, the pivotal movement may be related to the rotary movement morpheme in, e.g., BICYCLE: //ROTATE//. We have then a sequence, not a parallel set, of three morphemes, not phonemes or primes. The shift in level of analysis brings the sequential structure of the sign into focus. First the //FMO// occurs; then //POINT//, which is realized to agree in position and shape (the thumb is extended) with //FMO//. Finally, //ROTATE// is realized with a nodding action to agree with the prior environment. There is substantial temporal overlap: //POINT// and //FMO// are partly concurrent in execution and move toward agreement in location, orientation, and type of contact. The realization of these morphemes leads to an interleaved sequence of meaningless smaller units including the handshapes, /B,L/, and the movement notated as /ↄ/. Thus, just as analysis of spoken sequence leads to a view of speech as in some degree parallel in its execution, so an analysis of signed simultaneity leads to a view of signs as in some degree serial.

We should emphasize that, although the sequential structure of a sign has come into descriptive focus from a reanalysis of its posited prime set as a morpheme sequence, the description does not depend on that reanalysis. (Nor is this the place to propose the general recasting of ASL linguistic structure that this analysis implies.) Rather, the sequencing is entailed by the motoric dimensions themselves. In rapid signing, movement toward location must begin before complete formation of handshape, if location is not to be anomalous; and, if movement is not to be anomalous, handshape and location must be more or less fully established before sign-internal movement begins. In other words, a sequential structure seems intrinsic to sign formation, as a parallel structure is intrinsic to the spoken syllable.

CONCLUSION

We are led to the paradoxical conclusion that sign language draws on a degree of sequential organization to implement a parallel linguistic structure, while speech does precisely the reverse. But the paradox weakens if we see the two motoric modes as answers to the same communicative demand. The demand is for fluent discourse at a cognitively comfortable rate. The two languages then draw on the same linguistic competence and a common system of central motor control to meet this demand. Their solutions differ in emphasis because they deploy peripheral articulatory structures that differ in their degrees of freedom and that address different perceptual systems.

REFERENCES

Cooper, F. S. How is language conveyed by speech? In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, Mass.: M.I.T. Press, 1972

Fowler, C. A. Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 1979, *8*, 113-133.

Klima, E. Sound and its absence in the linguistic symbol. In J. Kavanagh & J. Cutting (Eds.), *The role of speech in language*. Cambridge, Mass.: M.I.T. Press, 1975.

Klima, E., & Bellugi, U. *The signs of language*. Cambridge, Mass.: Harvard University Press, 1979.

Öhman, S. Coarticulation in VCV utterance: spectrographic measures. *Journal of the Acoustical Society of America*, 1966, *39*, 151-168.

Stokoe, W., Casterline, D., & Croneberg, C. *A dictionary of American Sign Language*. Washington, D.C.: Gallaudet College Press, 1965.

Wang, W. S., & Fillmore, C. Intrinsic cues and consonant perception. *Journal of Speech and Hearing Research*, 1961, *4*, 130-136.

ARE MOVEMENTS PREPARED IN PARTS?
NOT UNDER COMPATIBLE [NATURAL] CONDITIONS*

David Goodman+ and J. A. Scott Kelso++

Abstract. This set of experiments is concerned with the specification of movement parameters hypothesized to be involved in the initiation of movement. Experiment 1 incorporated the precueing method developed by Rosenbaum (1980) in which a precue provided partial information of the upcoming movement prior to the stimulus to move. Under conditions in which precues were provided by letter symbols and stimuli were color-coded dots mapped to response keys, Rosenbaum (1980) found reaction times to be slower for the specification of arm than for direction, and both to be slower than the specification of extent. Under precue and stimulus conditions similar to those employed by Rosenbaum (1980), we obtained a similar trend. The three follow-up experiments extended these findings to more naturalized stimulus-response compatible conditions. We used a method in which precues and stimuli were directly specified through vision and mapped in a one-to-one manner with responses. In Experiment 2, although reaction times decreased as a function of the number of parameters precued, there were no systematic effects of precueing particular parameters. In Experiment 3, we incorporated an ambiguous precue that, while serving to reduce task uncertainty, failed to provide any specific information as to the arm, direction, or extent of the upcoming movement. However, initiation times did not systematically vary as a function of the type of parameter precued. Experiment 4 was a replication of Experiment 3, but there were no significant differences between specific or ambiguous precue conditions. In sum, only in Experiment 1 in which precues and stimuli involved complex cognitive transformations was there support for Rosenbaum's parameter specification model. When we employed

---

*Also in Journal of Experimental Psychology: General, 1980, 109, 475-495. A preliminary version of this paper entitled "Response selection versus feature selection in precued movements" was presented at the Ninth Canadian Psycho-motor Learning and Sport Psychology Symposium in Banff, Alberta, September 1977. A later version that included all the present experiments and titled "Are movements prepared in parts or as wholes?" was presented at the Psychonomic Society Annual Meeting, Phoenix, Arizona, November, 1979.

+Also University of Iowa.

++Also University of Connecticut.

Acknowledgment. This research was supported by grants AM25814 and NS13617 from the National Institute of Health and BRS grant RR05596. We thank G. Frekany for assistance in data collection for Experiment 4, B. Kerr for reviewing an earlier version, and G. Stelmach, S. Keele, and D. Larish for their helpful comments.

highly compatible conditions, we failed to obtain any tendency for movement parameters to be serially specified. We discuss grounds for suspecting the generality of parameter specification models and propose an alternative approach that is consonant with the dynamic characteristics of the motor control system.

One of the dominant facts to emerge in the area of movement control in the last decade is that complex sequences of behavior may be produced even when all information from the periphery is removed. Physiological evidence for the presence of endogenous neural networks in a variety of invertebrate phyla is now unassailable (e.g., Davis, 1976; Miles & Evarts, 1979; Stein, 1978). Moreover, it is well established that the isolated spinal cord of vertebrates possesses intrinsic functions capable of generating the basic flexion-extension pattern of locomotion (cf. Grillner, 1975; Shik & Orlovskii, 1976).

Direct efforts to extend these findings--often interpreted as evidence for "central programming"--to the coordination of human skilled movements have met with limited success. Reversible deafferentation methods have been employed in conjunction with various motor tasks (e.g., Laszlo, 1966), but interpretation of the resultant data is clouded by the co-occurence of sensory and motor impairment (Kelso, Stelmach, & Wanamaker, 1974) and the presence of residual sensation in nearby anatomical structures (Glencross & Oldfield, 1975).

An alternative approach, germane to the present article, is to use reaction time (or more properly, initiation time; Kerr, 1978) as an index of central motor preparation. The idea, first introduced by Henry and Rogers (1960), is simple. If a motor program is prepared in advance, the time to prepare it should be a reflection of the upcoming movement's complexity. In contrast, if no prior programming takes place, reaction time for simple and complex movements should not differ. There is a considerable body of data favoring the former proposition in both choice (cf. Klapp, 1977, for review) and simple reaction time paradigms (cf. Keele, 1980, for review).

Much of the recent work has been directed toward identifying the content of the basic programming unit; for example, the stress group (Sternberg, Wright, Knoll, & Monsell, 1980) or syllable (Klapp, Anderson, & Berian, 1973) in speech, or the type stroke in nonsense typing (Sternberg, Monsell, Knoll, & Wright, 1978). In addition, some investigators have related reaction time to various components of the upcoming movement such as extent and duration (cf. Kerr, 1978, for review). Little, however, is known about the actual construction of motor programs, an issue that Rosenbaum (1980) has addressed recently in some detail. Rosenbaum (1980) adopts an "information processing" view of motor programs in which the program is assumed to undergo progressive differentiation from some abstract, "nonmotoric" level to a "muscle-usable code." After cognitive decisions have been made, the role of the program, according to Rosenbaum, is to prescribe values on certain kinematic parameters (which he terms dimensions) that are under program control. A major question at this level of programming concerns how movement dimensions such as arm, direction, and extent are specified, and whether they follow any particular ordering rules. To investigate this issue, Rosenbaum introduced a movement precueing technique that took the following form: On a given trial a subject received prior information (via alphabetic letters) about all, some, or none of the values defining the upcoming response (e.g., RFX meant prepare a right-hand [R] forward [F] movement, the X providing no information about actual movement extent). Then, at the onset of the signal (a colored dot), the subject initiated the motor response. Assuming the subject used the precues effectively, initiation time should reflect the amount of time to program the value on the remaining, undefined parameter (in this example a short or long extent).

Using these procedures, Rosenbaum (1980) found that reaction time was shortest when extent was left to be selected, longer when a directional decision was required, and still longer when arm remained to be selected. Further, when two of three parameters had to be specified, reaction times were elevated overall and followed a pattern consonant with singly precued conditions. Although not ascribing a particular fixed order to the various parameters, Rosenbaum noted that arm, direction, and extent tended to be specified serially. The implications of these findings are potentially far reaching, and the technique itself (when combined with electrophysiological procedures) could afford new insights into the nature of movement initiation processes (cf. Requin, 1980, for a review of neurophysiological work on movement preparation).

Our first goal in this set of experiments, given the putative significance of Rosenbaum's (1980) results, was to replicate his major experiment (Experiment 1) in its entirety. This is not to imply that Rosenbaum did not perform a careful experiment and a thorough analysis, merely that we feel this often-ignored step constitutes sound practice. Overall, the pattern of results that emerges in our first experiment supports Rosenbaum's data quite well.

But if there is a flaw in the movement precueing technique as developed by Rosenbaum (1980), it is that the procedure itself is rather artificial. As indicated earlier, Rosenbaum used letters to precue the subject and previously learned color-coded labels as signals to respond. In our remaining experiments, we attempt to naturalize the precueing technique so that much less cognitive transformation (cf. Teichner & Krebs, 1974) is required. Our procedure was to precue the subject directly via vision and to map precues and stimuli with response buttons in a compatible manner. Thus, unlike Rosenbaum's procedure, which requires a color-to-position translation, our technique is referred to as direct because (a) it involves minimal stimulus coding activity and (b) the precue, stimulus, and response sets are in direct one-to-one correspondence. With these highly compatible procedures, which we feel are more representative of real-life motor skills, we demonstrate three basic findings: First, reaction times across precue conditions are considerably reduced over comparable conditions that require more stimulus-response translation time (e.g., our Experiment 1 and Rosenbaum's 1980, Experiment 1). Second, like Rosenbaum, reaction times are reduced as the amount of precue information increases. Third, but most important, within any particular precue condition the pattern of reaction times appears the same for all precued parameters. This last result, which shows no tendency for movement parameters to be serially ordered, persuades us of the need to reexamine the viability of "feature" specification models (Rosenbaum, 1980) especially when the geometric configuration of stimulus to response is naturalized and not artificially contrived.

EXPERIMENT 1

Experiment 1 was essentially a direct replication of Rosenbaum's (1980) first experiment with two additional modifications. Like Rosenbaum, we precued subjects by providing partial information about the upcoming movement and then required them to respond as quickly as possible to a stimulus by moving to the appropriate response key. Thus, some (or all) of the parameters of movement (e.g., arm and direction) could be prepared in advance, leaving only the remaining unknown parameter(s) (e.g., extent) to be specified. In addition, we incorporated two further experimental manipulations. First, two types of stimuli, a number or a color word, were used. Since a number to spatial location mapping requires fewer transformations than does a color word to spatial location (Teichner & Krebs, 1974), one might expect faster initiation times in the former case. Second, two precue durations, 3 and 5 sec, were employed to evaluate whether differential effects on parameter specification were due, in part, to incomplete precue processing.

Method

Subjects

Twenty-four right-handed persons between the ages of 18 and 30 yr. served as subjects. They were paid $5 for their services.

Apparatus

The experiment took place in a sound-insulated experimental chamber. The subject sat in an adjustable chair in front of a standard laboratory table 155 cm long, 66 cm wide, and 96 cm high. The reaction keys were mounted in a 46 cm x 31 cm Plexiglas base that was tilted at an angle of 20° to the horizontal. Two keys placed 21 cm apart and centered on the Plexiglas base served as the home keys for the left and right index fingers. Like Rosenbaum's (1980) configuration, eight target keys were situated so that two were directly above and two below each home key. The distance from the home keys to the near target was 3.5 cm and to the far target 7.0 cm. Home keys and reaction keys were standard keyboard switches (Cherry momentary contact switches) and required a 40-g operating force. The width of the response keys was equated for index of difficulty (Fitts, 1954; 1.3 cm diameter for near keys; 2.6 cm diameter for distant keys). A black piece of felt mounted above the response board prevented the subject from viewing the response keys but did not interfere with the response movements. A video computer terminal situated above and slightly behind the response board was used to display precues and stimuli. The precue consisted of capital letters displayed in the center of the video screen. Letters conveying arm information were R (right) and L (left). Letters conveying direction information were F (forward) and B (backward). Letters conveying extent information were C (close) and D (distant). Each precue consisted of three letters, and the letter X was used as a filler when the precue consisted of less than three informative letters. The reaction signal consisted of either a number (1-8) or a color word (e.g., RED). Each number or color word was mapped one-to-one to a response key. A Digital Equipment Corporation PDP 8/A computer was programmed to present the precues and the stimuli, as well as to time the initiation and movement times, and record them on a floppy disk for later off-line analysis.

Procedure

Each subject participated in a single experimental session lasting approximately 1 hr. and 20 min. Before testing began, subjects were given as much time as needed to familiarize themselves with the position of each response key and its unique mapping to a given stimulus. An initial block of 64 practice trials was performed for familiarization purposes. This was followed by two blocks of 128 trials, separated by a 3-min. rest period. The eight precue conditions (no precue; a single-parameter precue for arm, direction, and extent; a two parameter precue for arm and direction, arm and extent, and direction and extent; and a completely precued condition) were presented such that 16 trials of each precue condition occurred within each block. Each possible stimulus within each type of precue was presented equally often. This resulted in two stimulus response pairs to each of the eight response keys for each precue condition in each block.

The order of trials was randomized for each subject. The subjects were told the meaning of precues and were instructed to make use of them. Their task was to try to respond as quickly as possible without making errors. A trial sequence consisted of a precue display for 3 or 5 sec (depending on the condition), a fixed foreperiod of .5 sec, followed by the stimulus to move (either a number or color word, again dependent on experimental condition). The stimulus remained on the screen until the subject responded. Following the subject's response there was a 4-sec intertrial interval before the onset of the next precue.

Design

The first block of 64 practice trials was not included in any of the following analyses. There were, therefore, four responses to each of the eight response keys in each of the eight precue conditions, making a total of 256 trials. Trials in which the subject responded with the wrong hand, missed the response key, or hit the wrong response key were noted but excluded from the main data analysis. Furthermore, trials with reaction times greater than 2,000 msec (considered to be due to lack of attention) or less than 70 msec (considered to be due to anticipation of stimulus) and movement times greater than 600 msec were excluded.

Mean reaction time and mean movement time were computed for each combination of precue and response movement. Three types of analysis for each dependent measure were performed. The first analysis was conducted to determine the effect of the number of precued parameters. That is, the conditions of no precue, one precue (arm, direction, or extent), two precues (arm and direction, arm and extent, direction and extent), and the totally precued condition were treated as eight levels of precue condition in a six-way analysis of variance. Time of precue (3 or 5 sec) and type of stimulus presentation (number or color word) were between-group variables; precue condition (eight levels) and response movement (consisting of two levels of arm, direction, and extent) were repeated variables. The second analysis, to determine the effects of the different parameter(s) precued, was performed only on the three conditions in which one parameter was precued. Similarly, a third analysis, to determine the effect of the various combinations of two precued parameters, was performed only on the three conditions in which two parameters were precued. Error rates were examined in the same manner.

Results and Discussion

The analysis that follows will be discussed with respect to the three types of analysis performed. First we report reaction time, then movement time, and then errors.

Reaction Time Analysis

Full design. The mean reaction times for both the 3- and 5-sec precue display and for type of stimulus presentation (numbers and color-words) are shown as a function of precue condition in Figure 1.[1] This figure also displays the breakdown of response movement (arm--left/right, direction--forward/backward, extent--short/long) across all precue conditions. For reaction time there was a significant main effect of precue, $F(7, 140) = 190.1$, $p < .001$. Post hoc analysis of the main effect of precue using a Newman-Keuls test revealed that the completely precued condition was responded to fastest. The next fastest were those conditions in which only a single parameter remained to be specified (two parameters precued), followed by the singly precued condition, with the condition of no precue having the longest reaction time. These results appear to be accountable, at least in part, on the basis of uncertainty (Hick, 1952; Hyman, 1953). As the number of stimulus response alternatives was reduced (i.e., as more parameters were precued), there was a commensurate reduction in reaction time. Thus reaction time increased with the number of possible choices, whether these involved direction (Ells, 1973; Glencross, 1973; Kerr, 1976), extent (Glencross, 1973; Kerr, 1976), limb (Glencross, 1973), or any combination of the three parameters. This finding is consistent with Rosenbaum's (1980) finding that mean reaction times increased with the number of values to be specified after the reaction signal. Neither time of precue display (3 or 5 sec) nor type of precue (number or color word) was statistically significant ($Fs < 1$). However, there were some complex interactions involving both between- and within-subjects variables, the results of which are clarified in the following analyses.

One-precued parameter. To assess the main effects of interest, namely type of precue within the single precue condition (arm, direction or extent), four separate analyses of variance were carried out on the 3- and 5-sec number and color conditions. This procedure, basically a simple effects analysis, was carried out due to the complex interactions of the between-subjects variables (time of precue display and type of stimulus presentation) and some of the within-subjects variables. Precue type was crossed with response movement (two levels of arm, two levels of direction, and two levels of extent). In the 3-sec number condition, the main effect of precue type (arm, direction, or extent) failed to reach significance, $F(2, 10) = 2.08$, $p > .05$, nor were any interactions with precue type significant. With respect to response movements, the only significant result was in the extent condition, $F(1, 5) = 91.55$, $p < .01$, where shorter movements were initiated 34.1 msec slower than longer ones in spite of attempts to equate the movements in terms of index of difficulty. In the 5-sec number condition, there was no significant effect of precue type, $F(2, 10) = 3.68$, $p > .05$. None of the other main effects or interactions were significant.

The 3-sec color-word condition showed the same pattern of results as above with respect to precue type, $F(2, 10) = 3.36$, $p > .05$, but there was a

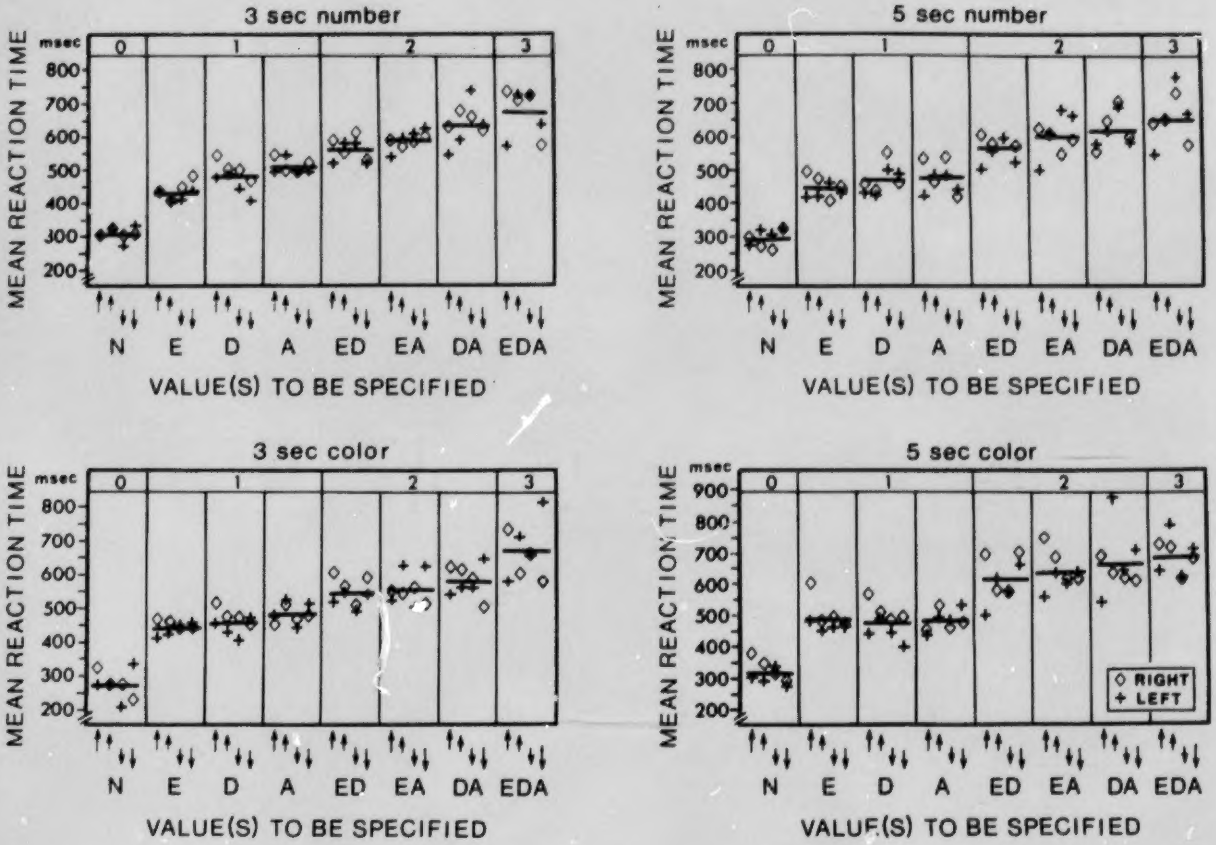


Figure 1. Mean reaction time for 3- and 5-sec precue displays and number and color-word stimulus presentations across the eight precue conditions. (In each condition, the overall mean is represented by a horizontal line. N=none; E=extent; D=direction; A=arm.)

three-way interaction involving response movements (Arm x Direction x Extent), $F(1, 5) = 21.7$, $p < .01$. For the left arm, short backward movements were initiated faster than short forward movements, whereas long forward movements were initiated faster than long backward movements. This effect was not present in right arm movements, a finding for which there is no ready explanation. Only in the 5-sec color-word condition was there an effect of precue, $F(2, 10) = 8.62$, $p < .01$. Post hoc analysis revealed that precueing arm resulted in faster initiation times than precueing movement extent but that neither precue type was reliably different from direction. A response movement interaction between direction and extent was also significant, $F(1, 5) = 8.02$, $p < .05$. Forward movements were initiated faster for longer extents, whereas backward movements were initiated faster for shorter extents.

Two precued parameters. An identical analysis to the one-precued parameter condition was carried out in the two-precue condition. In the 3-sec number condition, there was a main effect of precue type, $F(2, 10) = 5.92$, $p < .05$. Post hoc analysis revealed that precueing arm and direction (extent remaining to be specified) was faster than precueing direction and extent (arm remaining to be specified). In the 5-sec number condition, the main effect of precue was not significant, $F(2, 10) = 2.08$, $p > .05$, but precue did interact with direction, $F(2, 10) = 8.17$, $p < .01$. For backward movements, initiation time was faster when arm and direction were precued than when arm and extent were precued. But for forward movements, precueing arm and extent was significantly faster than precueing direction and extent. A response movement interaction between arm and direction was also evident, $F(1, 5) = 8.24$, $p < .05$: for the left arm, forward movement was initiated faster than backward movement, whereas for the right arm there were no directional differences.

In the 3-sec color-word condition, there was a significant precue effect, $F(2, 10) = 5.16$, $p < .05$. Further analysis revealed that movements were initiated faster when extent, rather than arm, remained to be specified (i.e., arm and direction versus direction and extent precued). No other effects were statistically significant. In the 5-sec color-word condition, there was no effect of precue type ($F < 1$). As in the 5-sec number condition, arm and direction interacted, $F(1, 5) = 7.12$, $p < .05$. But in this case, backward movements were initiated faster than forward movements only for the right arm.

Movement Time Analysis

A parallel breakdown of the experiment in terms of movement time to that provided in Figure 1 for reaction time is shown in Figure 2.

Full design. The initial analysis of the movement time data revealed that neither time of precue display (3 or 5 sec) nor type of stimulus presentation (number or color-word) were statistically significant (both Fs < 1). Nor were there any interactions involving these variables. There was a main effect of precue, $F(7, 140) = 7.19$, $p < .01$, which we explore in more detail in the following analysis.

One-precued parameter. In the single-precue condition, there were no effects of time of precue display or type of stimulus (Fs < 1). There was a main effect of precue, $F(2, 40) = 7.59$, $p < .01$. Precueing extent resulted in faster movements (21 msec) than precueing arm. Since this effect is in the

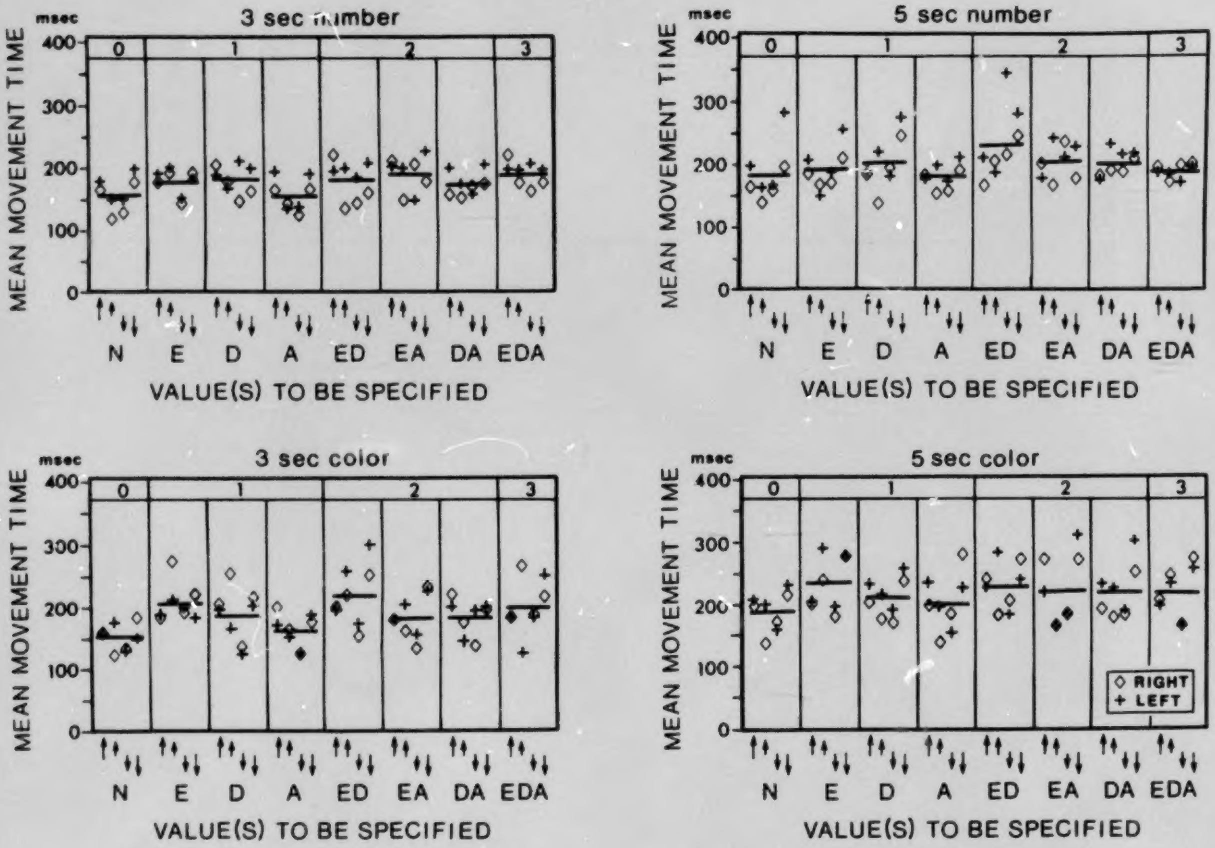


Figure 2. Mean movement time for 3- and 5-sec precue displays and number and color-word stimulus presentations across the eight precue conditions. (In each condition, the overall mean is represented by a horizontal line. N=none; E=extent; D=direction; A=arm.)

opposite direction to the trend evident in reaction time, there may be some type of trade-off between the two dependent variables. Movements of the right arm were made approximately 17 msec faster than those of the left, $F(1, 20) = 7.93$, $p < .05$. Movements to near targets were 27 msec faster on the average than movements to far targets, $F(1, 20) = 16.86$, $p < .01$, in spite of efforts to control for index of difficulty (Fitts, 1954). A three-way response movement interaction (Arm x Direction x Extent), $F(1, 20) = 4.60$, $p < .05$, indicated that the general finding of faster movement times for short movements was not present in left arm forward movements, which were actually slower for short than for long movements.

Two-precued parameters. The null findings of precue display time and stimulus display type were also apparent in the two-precue condition. Again, an effect of precue was present, $F(2, 40) = 8.94$, $p < .01$. Precueing extent and direction (arm to be specified) resulted in somewhat faster movement times (27 msec) than precueing arm and direction (extent to be specified). This finding poses a potential problem with the interpretation of the reaction time data because the two dependent variables go in opposite directions. That is, reaction time was longer in the 3-sec color and number conditions when arm rather than extent remained to be specified, but movement time was shorter in these conditions. This trade-off is not particularly suprising, since final extent can be determined after the movement has been initiated, whereas determination of arm must occur before movement initiation or an error occurs. As in single-precue conditions, short movements were carried out faster than long movements (29 msec on the average), $F(1, 20) = 19.81$, $p < .01$. The two-way response movement interaction between extent and direction, $F(1, 20) = 15.29$, $p < .01$, revealed this difference to be greater in backward than in forward movements.

Error-Rate Analysis

The error-rate data, differentiated by error type, are presented as a function of precue condition in Table 1. Although the error rate, averaged across precue duration and stimulus type, ranged from 3% to 11.2%, the no-precue condition (8.6%) and the totally precued condition (10.7%) were well within these ranges, suggesting that error rate, at least in this experiment, bore no particular relationship to stimulus-response uncertainty. Analysis of variance on each Precue Display Time (3 or 5 sec) by Stimulus Type (number or color word) combination revealed a main effect of precue only in the 3-sec color condition, $F(2, 10) = 4.76$, $p < .05$. Precueing extent (direction and arm to be specified) resulted in significantly more errors than precueing arm (extent and direction to be specified). This effect, however, does not change the interpretation of reaction time, as the error rate was lowest in the condition with the fastest reaction time.

In the two-precue condition, only the 3-sec number condition provided evidence for an effect of precue type, $F(2, 10) = 16.04$, $p < .01$. The error rate when extent and direction were precued was greater than that of the other two precue conditions. As in the single-precue condition, the directionality of the errors as a function of precue type followed the reaction time analysis.

Table 1

Percentage Error Rate Categorized by Error Type as a Function of Precue Conditions and Stimulus Presentation Type: Experiment 1

	Parameters to be specified							
Type of error	N	E	D	A	ED	EA	DA	EDA
3-sec number								
Anticipation[a]	2.6	.0	.5	7.8	1.0	6.3	5.2	6.3
Inattentiveness[b]	2.1	1.0	1.6	2.1	3.1	1.6	4.7	3.6
Response[c]	7.3	.0	.0	.0	.0	1.0	.0	.0
Total	12.0	1.0	2.1	9.9	4.1	8.9	9.9	9.9
5-sec number								
Anticipation	1.0	1.6	.5	5.2	2.1	10.4	8.3	6.3
Inattentiveness	4.2	4.2	3.1	1.6	6.8	2.6	3.1	2.6
Response	5.2	5.2	.0	2.1	.5	1.6	1.6	1.6
Total	10.4	11.0	3.6	8.9	9.4	14.6	13.0	10.5
3-sec color								
Anticipation	2.1	1.6	.5	6.8	.5	6.8	5.7	5.2
Inattentiveness	1.6	3.7	2.6	.5	3.7	1.0	3.1	2.6
Response	7.3	5.2	.0	.5	.0	.0	5.2	.0
Total	11.0	10.5	3.1	7.8	4.2	7.8	14.0	7.8
5-sec color								
Anticipation	3.1	.5	.5	6.8	2.1	10.9	8.3	3.7
Inattentiveness	.5	3.7	2.1	1.0	4.2	1.6	4.2	2.1
Response	4.7	.5	.0	.5	.0	1.0	.5	.5
Total	8.3	4.7	2.6	8.3	6.3	13.5	13.0	6.3

Note. N = none; E = extent; D = direction; A = arm. [a]Reaction times < 70 msec. [b]Reaction times > 2 sec. [c]Initiated movement with wrong hand, struck wrong response key, or missed target altogether.

The findings of Experiment 1 are generally in support of the differential parameter specification hypothesis (Rosenbaum, 1980), although the effects observed in our experiment are not always statistically reliable. For example, in the conditions in which two parameters were precued, only in the 3-sec number and 3-sec color condition were there statistical effects of precue type on reaction time. Similarly, in the conditions in which one parameter was precued, only the 5-sec color condition provided any statistical evidence for differential specification times. But when we compare our reaction time data and those of Rosenbaum, there is considerable similarity in the two sets of data (see Table 2). The inequality $B_A > B_D > B_D$, where these terms represent value specification times for arm, direction and extent, respectively, seems to hold in seven of the eight Precue Display Time by Stimulus Type conditions.

Table 2

Comparison of Reaction Times (in msec) in the Four Conditions of Experiment 1 and Rosenbaum's Experiment 1

Condition	Reaction time		
	One parameter precued		
	A	D	E
3-sec number	559	588	634
5-sec number	562	598	616
3-sec color	540	551	575
5-sec color	613	634	660
Rosenbaum[a]	537	565	591
	Two parameters precued		
	A and D	A and E	D and E
3-sec number	431	477	512
5-sec number	441	465	469
3-sec color	442	457	478
5-sec color	486	478	481
Rosenbaum[a]	434	461	489

Note. A = arm; D = direction; E = extent.
[a]From Rosenbaum (1980).

Some caution is warranted, however, in interpreting this trend completely in terms of parameter specification, at least prior to movement initiation. There was some evidence in the movement time data that extent decisions were actually made after the limb had begun to move. Rosenbaum (1980) observed a similar effect in his movement time data, and, clearly, kinematic information about movement trajectories would help clarify the issue. In addition, the magnitude of precue effects in our experiment diminished as precue display time was increased from 3 to 5. Interestingly, Rosenbaum (1980, Footnote 5) mentions an informal study indicating the same result but offers no rationalization for it. Perhaps the most realistic, though speculative, possibility is that the subject can make maximum use of the time to process precues: With additional time the need to employ a parameter specification strategy may be less crucial. On the other hand, and equally speculative, the expectancy state brought about by precueing the subject may have only a brief duration, after which the subject ceases to prepare individual response parameters. Why such a hypothetical state should extend to 3 but not 5 sec is somewhat mysterious.

Whatever the case, there is little doubt that the experimental situation created by Rosenbaum (1980) and by us in Experiment 1 is far removed from anything that would represent real-life movement control. Although there is little argument that animals and humans can effectively use prior information about upcoming movements of limbs (e.g., Kelso, Pruitt, & Goodman, 1978) and eyes (e.g., Bizzi, 1974) to control them effectively, it is rare indeed for such prior information to take the form of letter precues. Even less often (except possibly in psychological experiments) does an individual have to make color transformations to produce a movement. On the other hand, the extensive experiments of Simon and colleagues (e.g., Simon, 1969; Simon & Rudell, 1967) show that initiation and movement time performance improves considerably when the stimuli exploit "natural" response tendencies of subjects. The possibility arises therefore that the experimental arrangement employed by Rosenbaum and ourselves may be so far removed from reality that the data obtained may be quite irrelevant to the phenomenon of interest, namely, the parameterization of motor programs.

Even if one is suspicious about the need for ecological validity (which we believe is well motivated here, see Neisser, 1976, chap. 3 for discussion), Rosenbaum's results, which receive reasonable support in our Experiment 1, would be much stronger if obtained under more natural conditions. One way to examine this issue is to link spatially precues and stimuli more directly to responses (via vision) and thus reduce the number of cognitive transformations required. Recently, Lee (1980) has presented evidence from a wide variety of activities--preserving balance in a "swinging room," catching, hitting, driving a car--along with a detailed mathematical analysis of optical flow, demonstrating the intricate and nonarbitrary relationship between vision and the motor system. This coupling can also be well motivated at several different levels of neural processing (cf. Arbib, 1980, for review). In the experiments to follow, therefore, we mapped precues and stimuli to required responses in a highly compatible way. Thus, subjects received prior information about the parameters of upcoming movement via vision, and visual stimuli (not color-coded dots or names) specified the appropriate responses. There was then an attempt to maximize differential parameter specification by visual means and instructions to subjects about how to use this information effectively. If Rosenbaum is correct, that is, that his data speak to the "programming of movement" after nonmotoric decisions have been made, there is no a priori reason to expect the hypothesized differential parameterization effects obtained under these rather contrived conditions to be eliminated under more natural conditions.

EXPERIMENT 2

Method

Subjects

The subjects were 10 right-handed adults who were not paid for their services.

Apparatus

The apparatus was similar to that employed in Experiment 1, with one major modification in the way precues and stimuli were displayed to the subject. The video computer terminal was replaced by a display board (for precue and stimulus presentation), which consisted of a 21-cm x 41-cm Plexiglas board mounted vertically at eye level. Eight red light-emitting diodes were mounted in the same configuration as the response board. A ninth light-emitting diode mounted above the eight precue diodes was used to indicate that the display was a precue display rather than a stimulus to move. The same diodes served as the stimulus lights. A Digital Equipment Corporation PDP 8/A computer was programmed to present the precues and the stimuli, as well as to time initiation and movement times, and record them on floppy disk for later off-line analysis.

Procedure

Each subject participated in a single experimental session lasting approximately 1 hr. and 40 min. Within this session there were four blocks of 128 trials, each consisting of a randomly presented precue followed by a stimulus to respond, in the same trial sequence as in the previous experiment.

A single light-emitting diode on the display board was activated to precue a subject completely on all parameters. To precue a subject on a single parameter, four diodes were turned on. For instance, to precue the left arm, the four lights on the left appeared. Similarly, to precue a long extent, the outermost lights were activated. Thus there were two alternative ways that each of the three singly precued parameters could be signaled. Precueing two parameters simply involved turning on the diodes formed by the intersection of the two sets of individually precued parameters: To precue arm and direction, for instance, the left or right lights indicating a forward or backward direction were turned on. There were thus four different ways to present each condition in which two parameters were precued.

The order of trials was randomized for each subject. As in Experiment 1, the subjects were told the meaning of the precues and instructed to make use of them in order to respond as quickly as possible without making errors. A trial sequence consisted of a precue in which the appropriate light diodes were activated for 3 sec, a variable foreperiod randomly selected from a uniform distibution of .5 to 1.5 sec, followed by the stimulus to move. The stimulus light remained on until the subject responded. After the subject's response there was a 4 sec intertrial interval before the onset of the next precue.

Design

The first block of 128 trials was considered practice and was not included in the analysis. There were therefore six responses to each of the eight response keys in each of the eight precue conditions, making a total of 384 trials. Trials in which the subject responded with the wrong hand, missed the response key, or hit the wrong response key were noted and analyzed separately as errors. In addition, trials with reaction times greater than 600 msec or less than 70 msec and movement times greater than 600 msec were excluded for the same reasons as before.

A within-subjects design was used with all 10 subjects performing the same number of responses in each precue condition to each response key. From the six trials resulting from each combination of precue and response movement, a mean reaction time and movement time was computed. As in the previous experiments, three separate analyses of variance were performed on each of the dependent variables. The first was an overall analysis of all precue conditions. The second and third dealt with the single and two precue conditions, respectively. As in Experiment 1, precue condition was crossed with response movement, which consisted of two levels of arm, two levels of direction, and two levels of extent, resulting in a four-way repeated measures analysis of variance. In addition, within-subject correlation coefficients were computed between reaction time and movement time (over the 384 trials per subject), and errors were analyzed and tabulated.

Results and Discussion

Reaction Time Analysis

Full design. The mean reaction times are shown for each precue condition collapsed over response movement in Figure 3. For reaction times there was a significant main effect of precue, $F(7, 63) = 52.16$, $p < .01$. Post hoc analysis using a Neuman-Keuls procedure indicated that the completely precued condition was the fastest. The next fastest were those precue conditions in which two parameters were precued, followed by single precue and no precue conditions. This result replicates those of the first experiment as well as Rosenbaum (1980, Experiment 1) in which reaction times increased as a function of stimulus-response uncertainty.

One-precued parameter. In the single-precue condition, precue type was not significant, $F(2, 18) = 3.04$, $p > .05$, nor were any main effects of response movement (arm, direction, or extent) significant. Precue type did, however, interact with extent of movement, $F(2, 18) = 4.09$, $p < .05$. Post hoc analysis revealed that for short movements, precueing arm (specification of direction and extent required) resulted in slower initiation time than either precueing extent (Mean diff. = 18.1 msec) or direction (Mean diff. = 19.3 msec). In contrast, for long movements, precueing extent resulted in slower

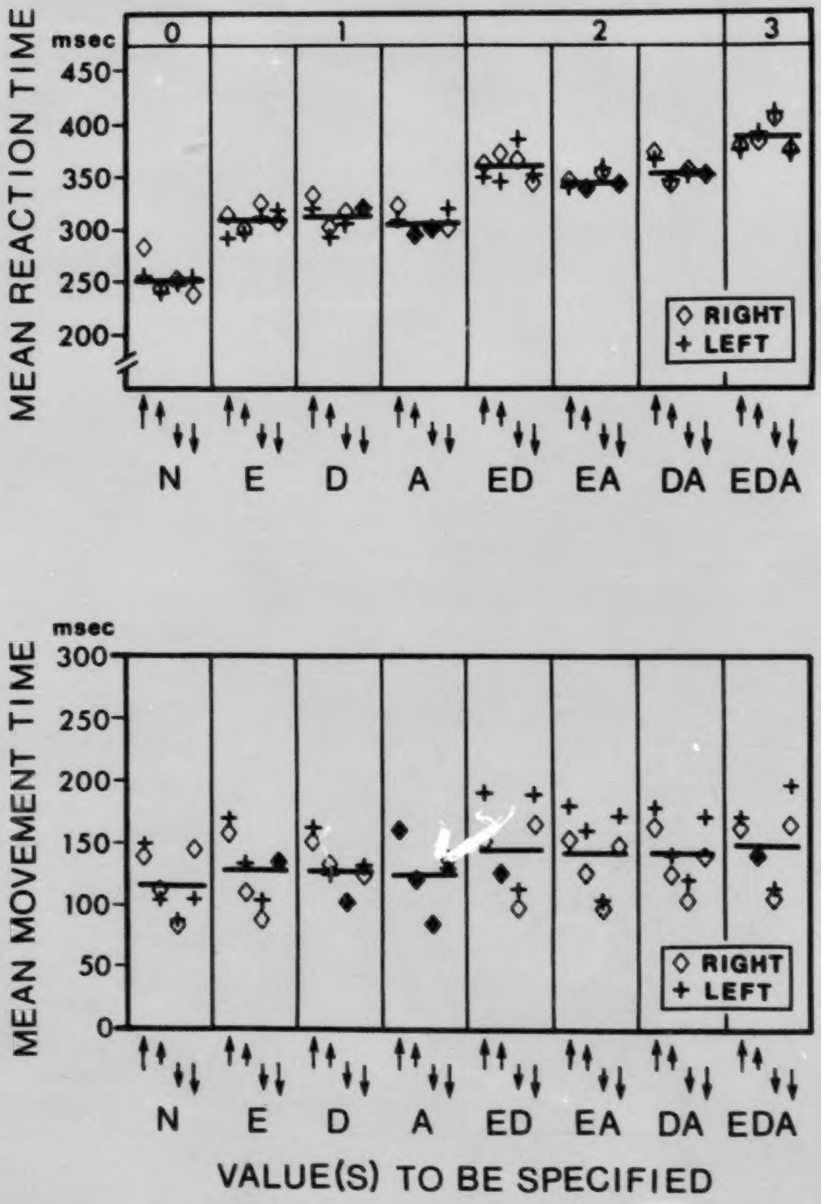


Figure 3. Reaction time and movement time for each precue condition in Experiment 2. (In each condition, the overall mean is represented by a horizontal line. N=none; E=extent; D=direction; A=arm.)

initiation times than precueing direction (Mean diff. = 17.0 msec). This particular interaction is troublesome for a model that predicts a fixed inequality of value specification times. That one inequality ($[B_A + B_D] < [B_D + B_E]$) should hold for short movements while another ($[B_A + B_E] < [B_A + B_D]$) should hold for long movements is less than parsimonious. Direction and extent of response movement also interacted in the singly precued condition, $F(1, 9) = 9.08$, $p < .05$. Forward movements were initiated faster to far than to near targets and backward movements were initiated faster for near than to far targets.

Two-precued parameters. In the two-precue condition (one parameter remaining to be specified), there was again no main effect of precue, $F(2, 18) = 1.79$, $p > .05$. However, as in the single precue condition, precue and extent of movement interacted, $F(2, 18) = 5.26$, $p < .05$. Post hoc analysis revealed that only in the longer movements was there a difference in reaction time based on type of precue; precueing arm and extent resulted in longer initiation times than precueing direction and arm. No other effects were statistically significant.

Movement Time Analysis

The mean movement times are shown for each precue condition collapsed over response movement in Figure 3.

Full design. The initial movement time analysis revealed a main effect of precue, $F(7, 63) = 8.20$, $p < .01$, which followed the same trend as the reaction time analysis with respect to number of precued parameters. When no parameters were precued, movement times were slowest, next slowest were the single precue conditions, followed by the two-parameter precued conditions. The totally precued condition exhibited fastest movement times. This finding lends support to those of Kerr (1976) and Fitts and Peterson (1964), where movement times were found to be slower as a function of either extent or directional uncertainty. More important, movement times follow the obtained reaction time pattern thus providing no evidence for a reaction time-movement time trade-off.

One-precued parameter. In the single-precue condition, there was no effect of precue, $F(1, 9) < 1$. Right-arm movements were performed approximately 20 msec faster than left, $F(1, 9) = 24.3$, $p < .01$. In addition, short movements were performed an average of 48 msec faster than long movements, $F(1, 9) = 76.8$, $p < .01$.

Two-precued parameters. The analysis of the two-precue condition revealed similar results to those reported in the one-precued parameter condition. No effect of precue was found, $F(2, 18) < 1$. Forward movements were approximately 26 msec slower than backward movements, $F(1, 9) = 6.98$, $p < .05$. Also, short movements were performed faster than long movements (Mean Diff. = 42 msec), $F(1, 9) = 81.32$, $p < .01$. This is consistent with Fitts's law (Fitts, 1954), where movement time increases as a function of distance when target size is held constant.

Error Rate Analysis

The error rate data are presented in Table 3. The average error rate across precue conditions was 8.4%, with the highest rate in the no-precue condition (13.5%). Error rates for individual subjects ranged from a low of 1.8% to a high of 14.0%. In the single precue condition, there was no effect of precue type on error rate, $F(2, 18) < 1$. However, more errors were made in movements to far targets ([illegible]7%) than to near targets (5.7%), $F(1, 9) = 8.45$, $p < .05$. There were no statistically significant results in the two precue condition, (Fs < 1).

Table 3

Percentage Error Rate Categorized by Error Type
for Each Precue Condition:
Experiment 2

	Parameter(s) to be specified							
Type of error	N	E	D	A	ED	EA	DA	EDA
Anticipation^a	5.4	2.9	4.4	3.9	3.8	3.3	3.9	5.0
Inattentiveness^b	2.5	2.3	1.7	2.5	1.5	3.1	1.7	5.6
Response^c	5.5	1.5	1.0	1.5	1.9	1.3	1.5	2.9
Total	10.4	6.7	7.1	7.9	7.1	7.7	7.1	13.5

Note. N = none; E = extent; D = direction; A = arm. ^{a}Reaction times < 70 msec. ^{b}Reaction times > 600 msec. ^{c}Initiated movement with wrong hand, struck wrong response key, or missed target altogether.

The within-subject correlation analysis revealed movement times to be largely independent of reaction times; all subjects' correlation values were less than ±.2.

The present results appear, as in Experiment 1, to be accountable to a large degree on the basis of uncertainty. As the number of stimulus-response alternatives was reduced (more parameters precued), there was a commensurate reduction in reaction time. Once again, reaction time increased with the number of possible choices of direction, the number of extent alternatives, and limb uncertainty. But unlike Rosenbaum (1980) and our Experiment 1, there were no systematic effects on reaction time within a particular precue condition. Rather, it appears that directly given precues allow the subject to eliminate particular stimulus-response alternatives and prepare those remaining in a more holistic manner. For example, in a situation in which two parameters are precued, the subject may prepare the two remaining responses (regardless of particular parameter) and simply choose between them when the stimulus light appears.

The foregoing "response selection" notion was examined by Rosenbaum (1980, Experiment 3). By identifying a response set (two or four choice) and instructing subjects to prepare multiple movements, Rosenbaum obtained similar findings to those reported here. But Rosenbaum's Experiment 3 bears little resemblance to the present experiment and is not particularly relevant to the claim we are making. First, in his Experiment 3 Rosenbaum used a color dot display and required subjects to learn a color dot to response-key mapping. In contrast, we used a directly compatible precue stimulus response mapping. Second, Rosenbaum actually instructed subjects to prepare multiple responses: We did not. Third, Rosenbaum used a precue display lasting 5 sec: We used a 3-sec precue display that our subjects, unlike Rosenbaum's (see Footnote 4 of Rosenbaum, 1980), had little difficulty identifying. We and Rosenbaum (1980, Footnote 5) have already shown that differential parameter specification effects are reduced or eliminated when precue display time is increased to 5 sec. The lack of evidence for such a process in Rosenbaum's Experiment 3 is therefore hardly suprising.

The results of the present experiment are more likely a reflection of the lack of robustness of the parameter specification model. Naturalizing the experimental situation appears to reduce parameter specification effects and may challenge their significance in the first place. Before rejecting the model, however, it is possible that individual parameters are specified, but that specification time is the same irrespective of the particular parameter involved (we will refer to this special case as nondifferential parameter specification). If this were the case, then two outcomes are predicted: First, reaction times should be similar when comparing conditions with the same number of parameters precued, and second, an increase in the number of parameters remaining to be specified should be accompanied by a corresponding increase in reaction time. Unfortunately, the same predictions follow from a response selection notion, and the data from Experiment 2 cannot discriminate between the two. This led us to the third experiment, whose purpose was to further enhance the likelihood of subjects using a parameter specification process as well as attempt to discriminate between parameter specification (differential or nondifferential) and response selection.

EXPERIMENT 3

Three major changes in procedure were incorporated into Experiment 3 to encourage parameter specification. First, trials were blocked on the type of parameter(s) precued. Thus, all trials within a single block involved precueing the same two parameters (e.g., extent and direction) such that a choice had to be made on the single remaining parameter (e.g., arm). Second, the subject was instructed to vocalize the information provided by the precue (e.g., forward, long) and to prepare those parameters. The third change was in the experimenter's role. Whereas in Experiment 1 and 2 the experimenter simply monitored the computer controlled experiment, in Experiment 3 the subject was verbally encouraged to prepare the response and respond as fast as possible. Verbal encouragement has been shown by Klapp, Wyatt, and Lingo (1974) to enhance preparation and facilitate the production of faster reaction times.

To investigate the hypothetical distinction between nondifferential parameter specification and response selection, a further condition was added in which the precue was rendered ambiguous. In this condition, the precue did not specify any particular parameter, but rather provided two stimulus response alternatives that differed in all three parameters. For example, consider a situation in which the visual precue specified a left forward movement to the far key and a right backward movement to the near key. Here parameter specification as envisaged by Rosenbaum (1980) would not be possible. On the other hand, even a nondifferential parameter specification model would predict reaction time differences between an ambiguously precued condition and a condition in which specific parameters were precued. But if the underlying process under compatible conditions involves response selection, reaction time should be the same across all situations in which there are two alternatives.

Method

Subjects

Eight right-handed adults who did not participate in either of the previous experiments served as unpaid subjects.

Apparatus

The apparatus was the same as that employed in Experiment 2. As in the first and second experiment, precue and stimulus presentation were computer controlled, with the response data collected and written out on floppy disk.

Procedure

Each subject participated in a single experimental session lasting approximately 40 min. Within this session there were four blocks of 40 precued trials followed by a stimulus to respond. Each trial consisted of a 3-sec precue display, during which the subject was required to announce the partial information conveyed by the precue. A 1/2-sec delay followed and preceded the stimulus to move. The intertrial interval was 3 sec. Within a single block the same two parameters were always precued, although in different manners. For instance, arm and direction could be signaled by precueing left-arm forward or backward movement and right-arm forward or backward movement. In each case the precue allowed the subject to partially prepare the type of movement specified, thus leaving the remaining parameter to be selected (extent in this case) when the stimulus occurred. Each combination of two precued parameters accounted for three of the experimental conditions. The fourth condition was designed so as not to precue any specific parameter, although leaving the same number of alternatives as the other conditions. For example, a left-arm forward movement to the far response key was paired with a right-arm backward movement to the near key.

Each possible stimulus was presented equally often within each precue condition. This resulted in five stimulus-response pairs to each of the eight response keys in each block. The order of precue conditions was counterbalanced. The subjects were given an initial period of time in which to become accustomed to the response movements by moving to each response key in succession for a total of five times. As in Experiments 1 and 2, there was no visual feedback from response movements. After the period of familiarization, subjects were advised as to the meaning of the precue display. At the start of each block, explicit instructions were given stressing the requirement to prepare the movement so that only the remaining parameter would have to be selected. Furthermore, each alternative precue within the upcoming condition was explained and demonstrated to the subject. After the first eight trials within each block, there was a short pause in which the experimenter informed the subject that he/she was going too slow (regardless of the actual speed of response). Again, preparation of response parameters was encouraged. After Trials 16, 24, and 32, the subjects were once again reminded of the importance of preparing the parameters prior to the response signal. The first eight trials within each block were considered practice trials and were excluded from the analysis. Trials in which the subject responded with the wrong hand, missed the response key, or hit the wrong response key were noted but excluded from the data analysis, as were trials in which reaction times or movement times were outside the ranges used in Experiment 2.

Design

A within-subjects design was used with all eight subjects performing the same number of choice reaction times in each precue condition to each response key. From the 4 trials resulting from each different response movement in each condition, mean reaction time and movement time were computed, which then served as the dependent variables in a 4 (precue) x 2 (arm) x 2 (direction) x 2 (extent) repeated measures analysis of variance. In addition, the error rate was analyzed in the same manner. A within-subjects correlation (for each block of 32 trials) between reaction time and movement time was computed.

Results and Discussion

Reaction Time Analysis

Mean reaction times are shown for each precue condition in Figure 4. The main effect of interest, type of precue condition, failed to reach significance, $F(3, 21) = 2.69$, $p > .05$. The only statistically significant result was for arm, $F(1, 7) = 6.36$, $p < .05$. Left-arm movements were initiated approximately 21 msec faster than right-arm movements. The null findings of precue condition are consistent with the null findings obtained for precue type in Experiment 2, since each precue condition had the same amount of uncertainty. Again, there was no evidence to suggest that response parameters were differentially specified. The finding that the ambiguously precued condition was not significantly different from the other precue conditions is not consistent with a general parameter specification process. Rather, each precue condition contained the same amount of uncertainty and thus appeared to exhibit the same reaction times. Reaction times in this experiment were somewhat faster (28 msec on the average) than comparable conditions in Experiment 2, suggesting that either verbal encouragement or the blocking of trials or both were effective means of speeding responses.

Movement Time Analysis

Mean movement times are shown for each precue condition in Figure 4. Analysis revealed that short movements were performed an average of 50 msec

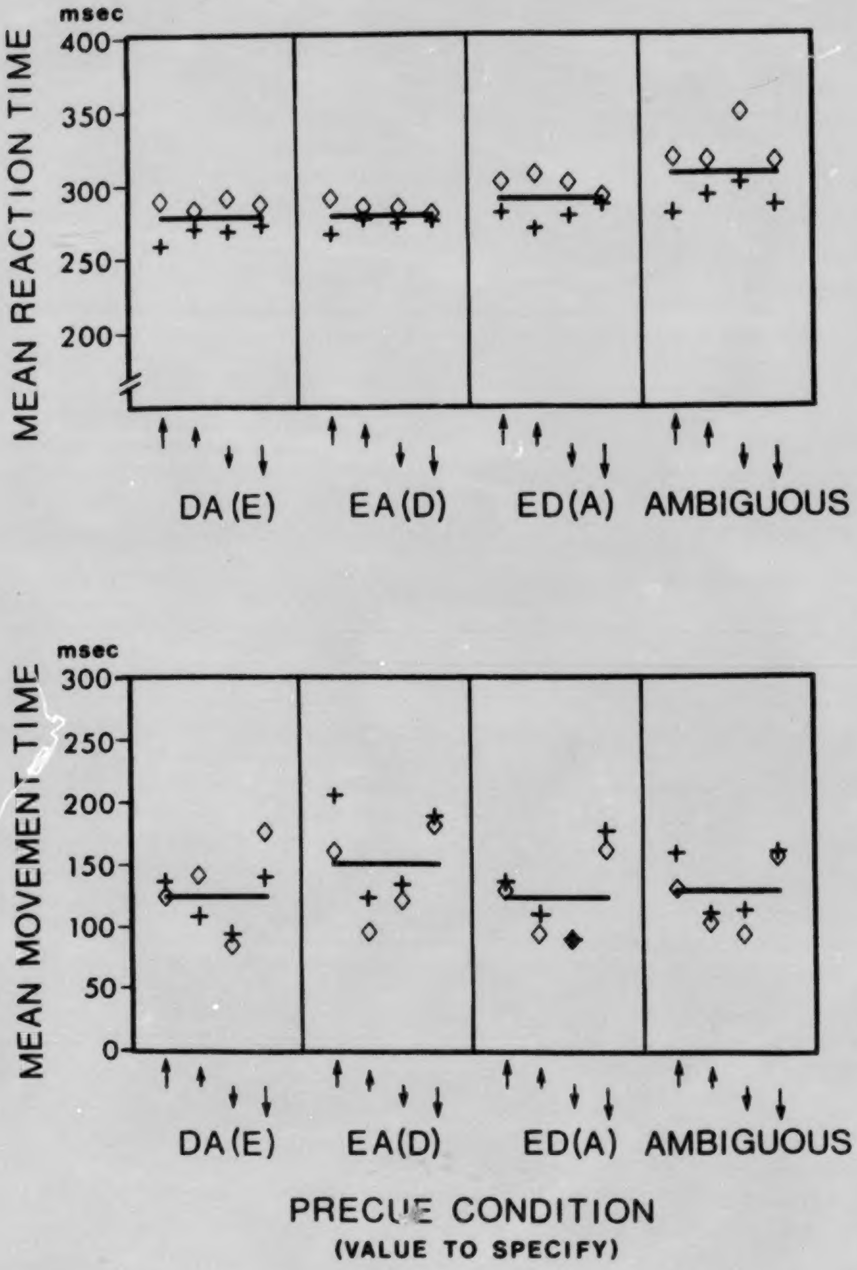


Figure 4. Reaction time and movement time for each precue condition in Experiment 3. (In each condition, the overall mean is represented by a horizontal line. D=direction; A=arm; E=extent.)

faster than long movements, $F(1, 7) = 15.14$, $p < .01$. The only other statistically significant finding was the Precue x Arm interaction, $F(3, 21) = 3.19$, $p < .05$. When arm and direction were precued, movement time was shorter for the left arm, whereas in the other precue conditions, movement times were shorter for the right arm.

Error-Rate Analysis

The percentage error rate for each precue condition is shown in Table 4. The analysis of the error rates indicated no differences across precue conditions, $F(3, 21) \sim 1$, nor were any other effects significant. The average error rate was 12.2%, ranging from a low of 9.8% when direction and arm were precued to a high of 15.6% when arm and extent were precued. The range for individual subjects spanned from 5.4% to 22.6%. The within-subject correlation analysis indicated that movement times and reaction times were virtually independent (all rs less than ±.28) as in Experiment 2.

Table 4

Percentage Error Rate Categorized by Error Type for Each Precue Condition: Experiment 3

Type of error	AD(E)[d]	AE(D)	DE(A)	Ambiguous (EDA)
	Precue condition			
Anticipation[a]	3.5	6.6	4.3	5.5
Inattentiveness[b]	2.0	3.1	2.7	2.0
Response[c]	4.3	5.9	5.5	3.4
Total	9.7	15.6	12.5	10.9

Note. A = arm; D = direction; E = extent.

[a] Reaction times < 70 msec. [b] Reaction times > 600 msec. [c] Initiated movement with wrong hand, struck wrong response key, or missed target altogether.

[d] Parameter(s) to be specified are in parentheses.

The present data are not particularly conducive to a parameter selection model, even one of the nondifferential kind. However, null effects must always be interpreted with caution, due to the possibility of Type II error. To counteract erroneous interpretation, we increased the number of subjects ($n = 24$) in a fourth experiment to increase the sensitivity of the experiment. In addition, six of the eight subjects in Experiment 3 indicated that verbalizing the upcoming movement seemed to interfere rather than aid planning of movement, so we excluded overt verbalization of the upcoming movements as well as experimenter encouragement to respond faster. Apart from these changes, the methods and procedures were identical to Experiment 3.

EXPERIMENT 4

Results and Discussion

Reaction Time Analysis

Mean reaction times are shown for each precue condition in Figure 5. As in Experiment 3, the main effect of type of precue failed to reach significance, $F(3, 69) = 2.43$, $p > .05$. However, there was a significant Precue x Extent interaction, $F(3, 69) = 4.74$, $p < .01$. For short movements the ambiguously precued condition resulted in the slowest initiation times overall, whereas in long movements, the condition in which direction remained to be specified (arm and extent precued) resulted in the slowest initiation times. With this exception, type of precue had no significant effect on reaction time. Indeed, the slowest initiation time (when direction remained to be specified) was only 14.4 msec slower than the condition with the fastest initiation time (when arm remained to be specified). Initiation times, on the average, were elevated approximately 20 msec beyond those obtained in Experiment 3, a result that may be due to removal of experimenter encouragement. Left-arm movements were initiated approximately 11 msec faster than right-arm movements, $F(1, 7) = 12.61$, $p < .01$, which replicates the left-arm advantage found in Experiment 3. Short movements were initiated faster in forward movements, whereas responses to far targets were initiated faster in backward movements, as indicated by the direction x extent interaction, $F(1, 23) = 28.90$, $p < .01$. As in the previous experiment, the reaction time data appear to provide little support for a general parameter selection process.

Movement Time Analysis

Mean movement times are shown for each precue condition in Figure 5. The movement time analysis revealed no effect of precue condition, $F(3, 69) \sim 1$, nor were any interactions with precue statistically significant. As in previous analyses, short movements were performed faster than long ones (Mean diff. = 56 msec), $F(1, 23) = 58.12$, $p < .01$. Like Experiment 3, forward movements were faster than backward movements (Mean diff. = 17 msec), $F(1, 23) = 12.31$, $p < .01$, and right-arm movements were made approximately 17 msec faster than left-arm movements, $F(1, 23) = 19.29$, $p < .01$.

Error Rate Analysis

The percentage error rate for each precue condition is shown in Table 5. The analysis of errors revealed no main effect of precue condition, $F(3, 69) = 2.40$, $p > .05$, whose average error rate was 6.2%. Forward movements had a higher error rate than backward movements, $F(1, 23) = 4.81$, $p < .05$, and long movements were more prone to error than short movements, $F(1, 23) = 27.87$, $p < .01$. An ordinal interaction between extent and direction, $F(1, 23) = 5.96$, $p < .05$, revealed the difference in error rates between forward and backward movements to be greater for longer movements. The within-subject correlational analysis indicated that movement times and reaction times were again relatively independent (all rs with one exception were less than $\pm .26$) as in the previous experiments.

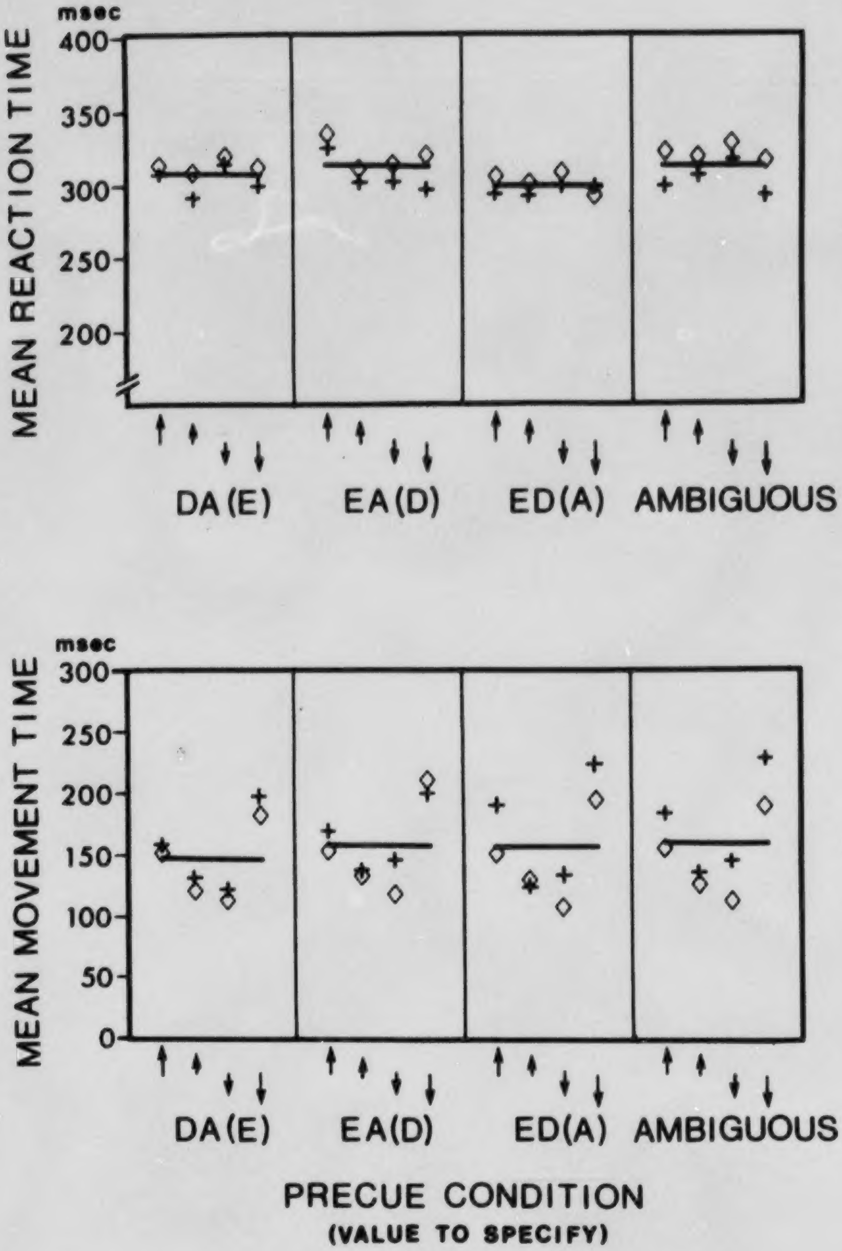


Figure 5. Reaction time and movement time for each precue condition in Experiment 4. (In each condition, the overall mean is represented by a horizontal line. D=direction; A=arm; E=extent.)

Table 5

Percentage Error Rate Categorized by Error Type for Each Precue Condition: Experiment 4

Type of error	AD(E)[d]	AE(D)	DE(A)	Ambiguous (EDA)
Anticipation[a]	2.2	2.7	1.6	3.4
Inattentiveness[b]	1.7	1.4	.7	2.1
Response[c]	2.9	1.7	3.6	3.5
Total	6.8	5.8	5.9	9.0

Note. A = arm; D = direction; E = extent.
[a] Reaction time < 70 msec. [b] Reaction time > 600 msec. [c] Initiated movement with wrong hand, struck wrong response key, or missed target altogether. [d] Parameter(s) to be specified are in parentheses.

GENERAL DISCUSSION

The present experiments were concerned with "programming" processes hypothesized to be involved in the initiation of simple movements. Our specific interest was whether the specification of movement parameters tended to proceed in a particular serial order as suggested by Rosenbaum (1980). The first experiment used the precueing method developed by Rosenbaum (1980) and was largely supportive of his main results. That is, there was indeed a definite tendency, admittedly not always statistically significant, for reaction times to be slower for the specification of arm than direction, and both to be slower than the specification of extent. In fact, there was some evidence in the movement time data to suggest that decisions about extent were actually made after the movement had been initiated, an effect also noted by Rosenbaum. Although this replication is heartening, the main thrust of the present article is directed toward extending these findings, if possible, to an experimental situation that bears a closer resemblance to the real-world task of controlling movement. More pointedly, the issue is one of evaluating whether the paradigm developed by Rosenbaum and employed in our Experiment 1 is really directed to the intended problem of interest, namely, the specification of motor program parameters after nonmotoric decisions have been made (Rosenbaum, 1980). Thus, subjects in Rosenbaum's main experiment and our Experiment 1 not only had to determine the meaning of letter precues but also had to translate a color-coded dot (name or number) into an appropriate response pattern. All this seems far removed from the skilled movement situation in which limb movements must be consonant with visually specified environmental changes.

In our follow-up experiments we employed a modification of Rosenbaum's (1980) method in which precues and stimuli were directly specified through vision. In the language of information processing and mental chronometry, we provided the subject with highly compatible stimulus response conditions. Thus, much less cognitive work is involved (or in Teichner & Krebs' 1974 analysis, fewer translational processes), a claim that receives strong support in the much faster reaction times observed in our Experiments 2-4, (see also Larish, 1980).2

In Experiment 2, although reaction times decreased as a function of the number of parameters precued, there were no systematic effects of precueing particular parameters.3 In Experiment 3, we incorporated a precue that, although serving to reduce task uncertainty, failed to provide any specific information as to the arm, direction, or extent of the upcoming movement. The parameter specification model predicts initiation time to be slower in this condition (termed ambiguous) than one in which some of the parameters of movements are known in advance. Such was not the case, however, as we again failed to detect movement initiation differences as a function of the type of precued parameter. Our reluctance to impute significance to null findings led us to replicate Experiment 3 with a larger sample. However, in a fourth experiment we again obtained null findings; there were no significant differences between specific or ambiguous precue conditions. In sum, of the four experiments we have performed, only in the one that used precues and stimuli of a quite complex kind (letters, color words, and numbers) did we find support for Rosenbaum's parameter specification model. When we employed highly compatible conditions, we failed to obtain any tendency for movement parameters to be serially ordered.

To the extent that compatible conditions are more natural for the subject (performance is certainly improved), we feel that some caution is warranted in adopting Rosenbaum's paradigm and generalizing his conclusions beyond the somewhat contrived situation in which the data were obtained. Note that we are not questioning the usefulness of precueing per se: This is an interesting innovation and may be very useful indeed as a tool to investigate the general nature of preparation (Kelso, in press). Our reservations speak to the specific precueing method and stimulus presentation employed by Rosenbaum (1980) and in our Experiment 1. Our suspicion, supported by the present data, is that this method has little to do with the parameterization of motor programs, at least at the motoric level that we and Rosenbaum are interested in. If the parameter specification model envisaged by Rosenbaum were a robust one, we would not have expected the ordering effects to wash out under more natural compatible conditions.

On hindsight there are grounds for questioning the viability of models of movement initiation positing (even tendencies in) serial ordering and partial preparation of motor programming parameters. For example, serial order notions run into a class of problems that mathematicians refer to as nondeterministic polynominal-time-complete (Lewis & Papdimetrios, 1978). In short, the only known algorithmic solution for such problems is one in which the execution time increases exponentially as a function of the number of variables to be regulated. Although only three parameters were investigated here, if one adopts the logical extension of this approach, more and more parameters must necessarily come into play as the task becomes increasingly more complex. This would necessarily result in an inordinate increase in programming time.

A further consideration with respect to parameter selection models is one raised by Kerr (1978). Task-defined parameters (such as arm, direction, and extent) may be quite different from the internal values that truly affect the motor control system. Thus, the parameters that experimenters define may not be considered singly or may not have one-to-one mappings in the motor control system. For instance, distance or extent of movement is not, as Keele (1980) points out, in the language of muscles, but instead is a consequence of the muscular forces that accelerate and decelerate the limb. From our perspective, the evaluation of programming effects on kinematic variables may be inappropriate: Kinematic measures are merely resultants of the system's dynamics.

Let us pursue briefly the dynamics perspective. Recent work in motor control strongly suggests that the natural physical properties inherent in neuromuscular systems (e.g., damping, stiffness) are exploited during movement. They are not merely the substrate on which central commands are laid down (cf. Bahill & Stark, 1979; Bizzi, Dev, Morasso, & Polit, 1978). For example, Polit and Bizzi (1978) have shown that the final position of the limb following reaching movements in monkeys is determined via the specification of stiffness and damping parameters that establish an equilibrium point between opposing pairs of muscles. Analogous experiments have been carried out in humans (Fel'dman, 1966; Kelso & Holt, 1980) and have led to models of single trajectory movements (such as those employed in these experiments) that possess the properties of homeomorphic oscillatory systems, the most specific being the mass spring (Kelso, 1977; Polit & Bizzi, 1978; Kelso, Holt, Kugler & Turvey, 1980). Hollerbach (1978) extended these findings by showing that cursive handwriting may be produced via coupled oscillations in the horizontal and vertical joints of the wrist-hand linkage. In Hollerbach's analysis, letters emerge from a constrained modulation of an underlying (dynamic) oscillatory process rather than a stringing together of individual motor programs. The consequence of the dynamics perspective, then, in contrast to one that views parameters as programmed for each individual movement, is that so-called complex movement behavior falls out as the modus operandi of a simple oscillatory pattern.

This view of coordination and control of movement as an emergent property of oscillator interactions contrasts sharply with a view of motor programs that prescribes parameters in whatever code is appropriate to get the correct muscles to fulfill the prescription (Rosenbaum, 1980). The latter assigns to the program a priori status in rationalizing motor behavior and in so doing ignores the fundamental problem for a motor control system; namely, how to regulate its internal degrees of freedom (Bernstein, 1967; Greene, 1972; Iberall & McCulloch, 1969; Turvey, 1977). In short, programming approaches, consonant with the computer metaphor, assign priority to the order grain of analysis and neglect entirely the relation grain (see Shaw & Turvey, in press, for a formal analysis of this issue). Programming languages (of computers and motor systems) are thus unidirectional and "imperative" (Steele & Sussman, 1978): in computers, command algorithms are separate from that which performs the computation just as the central program, in control theory and information processing approaches, is held conceptually distinct from the skeletomuscular apparatus that performs the movement.

We suspect that an adequate account of systemic movement behavior must, in the long run, include, as minimal requirements, a dynamic vocabulary for control (see above) and, relatedly, extend the explanation to the relational grain of analysis (cf. Gelfand, Gurfinkel, Tsetlin, & Shik, 1971; Greene, 1978; Boylls, 1975; Kelso et al., 1980; Kugler, Kelso, & Turvey, 1980; Shaw & Turvey, in press; Turvey, Shaw, & Mace, 1978). The latter promotes a search for the constraints that allow neuromuscluar variables to be regulated in a given motor activity. In fact, some progress has already been made in this regard. Nashner (1976), for example, has shown that over wide variations in upright posture brought about by ankle rotation, the ratios and sequencing of electromyographic activity in the muscles of the ankle, knee, and hip remain fixed. In handwriting, the timing of strokes remains fixed over changes in letter size and increases in friction between pen and surface (cf. Wing, 1978). Similarly, the timing relations of the upper limbs during the performance of a task involving different spatial demands remains invariant over changes in the magnitude of force produced by each limb (Kelso, Southard, & Goodman, 1979). In sum, the fixed proportioning of activity throughout a collection of muscles and the maintenance of timing relationships is a consequence of the constraints on the system. It is not, we should emphasize, that movements are caused by constraints, rather it is that some movements are excluded by them. This analysis leads us to suspect that an act is not the outcome of a collection of parameterizations dispersed in time but rather may be centrally or peripherally manipulated as a holistic structure.

REFERENCE NOTE

1. Frekany, G. A., Kelso, J. A. S., & Goodman, D. Processing demands of response programming. Unpublished manuscript, Motor Behavior Laboratory, University of Iowa, 1978.

REFERENCES

Arbib, M. A. Perceptual structures and distributed motor control. In V. B. Brooks (Ed.), Handbook of physiology: Vol 3, Motor control. Bethesda, Md.: American Physiological Society, 1980.

Bahill, A. T., & Stark, L. The trajectories of saccadic eye movements. Scientific American, 1979, 240, 108-117.

Bernstein, N. The coordination and regulation of movements. London: Pergamon Press, 1967.

Bizzi, E. The coordination of eye-hand movements. Scientific American, 1974, 231, 100-106.

Bizzi, E., Dev, P., Morasso, P., & Polit, A. Effect of load disturbances during centrally initiated movements. Journal of Neurophysiology, 1978, 41, 542-555.

Boylls, C. C. A theory of cerebellar function with applications to locomotion: II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat. (COINS Technical Report 76). Amherst, Mass.: University of Massachusetts, Department of Computer and Information Science, 1975.

Davis, W. J. Organizational concepts in the central motor networks of invertebrates. In R. H. Herman, S. Grillner, P. S. G. Stein, & D. G. Stuart (Eds.), *Advances in behavioral biology, Vol. 18: Neural control of locomotion*. New York: Plenum Press, 1976.

Ells, G. Analysis of temporal and attentional aspects of movement. *Journal of Experimental Psychology*, 1973, *99*, 10-21.

Fel'dman, A. G. Functional tuning of the nervous system during control of movement or maintenance of a steady posture-III. Mechanographic analysis of the execution by man of the simplest motor tasks. *Biophysics*, 1966, *11*, 766-775.

Fitts, P. M. The information complexity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 1954, *47*, 381-391.

Fitts, P. M., & Peterson, J. H. Informational capacity of discrete motor responses. *Journal of Experimental Psychology*, 1964, *67*, 103-112.

Gelfand, I. M., Gurfinkel, V. S., Tsetlin, M. L., & Shik, M. L. Some problems in the analysis of movements. In I. M. Gelfand (Ed.), *Models of the structural-functional organization of certain biological systems*. Cambridge, Mass.: MIT Press, 1971.

Glencross, D. J. Response complexity and the latency of different movement patterns. *Journal of Motor Behavior*, 1973, *5*, 95-104.

Glencross, D. J., & Oldfield, S. R. The use of ischemic nerve block procedure in the investigation of the sensory control of movements. *Biological Psychology*, 1975, *2*, 227-236.

Greene, P. H. Problems of organization of motor systems. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 2). New York: Academic Press, 1972.

Greene, P. H. *Organization of multivariable control tasks*. (Technical Report 78-12). Chicago: Institute of Technology, Computer Science Department, 1978.

Grillner, S. How detailed is the central pattern generator for locomotion? *Brain Research*, 1975, *88*, 367-371.

Henry, F. M., & Rogers, D. E. Increased response latency for complicated movements and a 'memory drum' theory of neuromotor reaction. *Research Quarterly*, 1960, *31*, 448-458.

Hick, W. E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 1952, *4*, 11-26.

Hollerbach, J. M. *A study of human motor control through analysis and synthesis of handwriting*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1978.

Hyman, R. Stimulus information as a determiner of reaction time. *Journal of Experimental Psychology*, 1953, *45*, 11-26.

Iberall, A. S., & McCulloch, W. S. The organizing principle of complex living systems. *Transactions of the American Society of Mechanical Engineers*, 1969, *91*, 290-294.

Keele, S. W. Behavioral analysis of motor control. In V. B. Brooks (Ed), *Handbook of physiology* (Vol. 3): *Motor control*. Bethesda, Md.: American Physiological Society, 1980.

Kelso, J. A. S. Motor control mechanisms underlying human movement reproduction. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 529-543.

Kelso, J. A. S. Remarks on the preparation of movement. In E. Donchin (Ed.), *Proceedings of the 1st Carmel Conference on Cognitive Neurophysiology*. Hillsdale, N.J.: Erlbaum, in press.

Kelso, J. A. S., & Holt, K. G. Exploring a vibratory system analysis of human movement production. *Journal of Neurophysiology*, 1980, *43*, 1183-1196.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach and J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland Publishing Company, 1980.

Kelso, J. A. S., Pruitt, J. H., & Goodman, D. The anticipatory control of movement. In G. C. Roberts & K. M. Newell (Eds.), *Psychology of motor behavior and sport*. Champaign, Ill.: Human Kinetics, 1978, 192-203.

Kelso, J. A. S., Southard, D. L., & Goodman, D. On the nature of human interlimb coordination. *Science*, 1979, *203*, 1029-1031.

Kelso, J. A. S., Stelmach, G. E., & Wanamaker, W. M. Behavioral and neurological parameters of the nerve compression block. *Journal of Motor Behavior*, 1974, *6*, 179-190.

Kerr, B. Decisions about movement direction and extent. *Journal of Human Movement Studies*, 1976, *3*, 199-213.

Kerr, B. Task factors that influence selection and preparation for voluntary movements. In G. E. Stelmach (Ed.), *Information processing in motor control and learning*. New York: Academic Press, 1978.

Klapp, S. T. Reaction time analysis of programmed control. In R. S. Hutton (Ed.), *Exercise and sport sciences reviews*. Santa Barbara, Calif.: Journal Publishing Affiliates, 1977.

Klapp, S. T., Anderson, W. G., & Berian, R. W. Implicit speech in reading reconsidered. *Journal of Experimental Psychology*, 1973, *100*, 368-374.

Klapp, S. T., Wyatt, E. P., & Lingo, W. M. Response programming in simple and choice reactions. *Journal of Motor Behavior*, 1974, *6*, 263-271.

Kugler, P. N., Kelso, J. A. S., & Turvey M. T. On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland Publishing Company, 1980.

Larish, D. *On the relationship between response organization processes and response programming*. Unpublished doctoral dissertation, University of Wisconsin, 1980.

Laszlo, J. I. The performance of a simple motor task with kinesthetic sense loss. *Quarterly Journal of Experimental Psychology*, 1966, *18*, 1-8.

Lee, D. N. Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland Publishing Company, 1980.

Lewis H. R., & Papdimetrios, C. H. The efficiency of algorithms. *Scientific American*, 1978, *238*, 96-109.

Miles, F. A., & Evarts, E. V. Concepts of motor organization. *Annual Reviews of Psychology*, 1979, *30*, 327-362.

Nashner, L. M. Adapting reflexes controlling the human posture. *Experimental Brain Research*, 1976, *26*, 59-72.

Neisser, U. *Cognition and reality*. San Francisco: Freeman, 1976.

Polit, A., & Bizzi, E. Processes controlling arm movements in monkeys. *Science*, 1978, *201*, 1235-1237.

Requin, J. Towards a psychobiology of preparation for action. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland Publishing Company, 1980.

Rosenbaum, D. A. Human movement initiation: Specification of arm, direction, and extent. *Journal of Experimental Psychology: General*, 1980, *109*, 444-474.

Shaw, R., & Turvey, M. T. Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy & T. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, N.J.: Erlbaum, in press.

Shik, M. L., & Orlovskii, G. N. Neurophysiology of locomotor automatism. *Physiological Review*, 1976, *56*, 465-501.

Simon, J. R. Reactions toward the source of stimulation. *Journal of Experimental Psychology*, 1969, *81*, 174-176.

Simon, J. R., & Rudell, A. P. Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 1967, *51*, 300-304.

Steele, G. L., & Sussman, G. J. *Constraints*. Artificial Intelligence Memo 502, Massachusetts Institute of Technology, November, 1978.

Stein, P. S. G. Motor systems, with specific reference to the control of locomotion. *Annual Review of Neuroscience*, 1978, *1*, 61-81.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning*. New York: Academic Press, 1978.

Sternberg, S., Wright, C. E., Knoll, R. L., & Monsell, S. Motor programs in rapid speech: Additional evidence. In R. A. Cole, (Ed.), *Perception and production of fluent speech*. Hillsdale, N.J.: Erlbaum, 1980.

Teichner, W. H., & Krebs, M. J. Laws of visual choice reaction time. *Psychological Review*, 1974, *81*, 75-98.

Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hillsdale, N.J.: Erlbaum, 1977.

Turvey, M. T., Shaw, R. E., & Mace, W. Issues in the theory of action. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum, 1978.

Wing, A. M. Response timing in handwriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning*. New York: Academic Press, 1978.

FOOTNOTES

[1] Note that on the ordinate of all figures we equate "value(s) to be specified" with "precue condition" for ease of interpretation and comparison with Rosenbaum's (1980) data.

[2] Larish (1980), in an independent study, also showed that transformation and translation processes (manipulated with various stimulus response configurations) were an important determiner of differential precueing effects.

[3] Frekany, Kelso, and Goodman (Note 1), in a study designed to evaluate the attentional demands of precues, had a built-in replication of Experiment 2. Results were virtually identical.

VELOPHARYNGEAL FUNCTION: A SPATIAL-TEMPORAL MODEL*

Fredericka Bell-Berti

I. INTRODUCTION

Speech sounds are produced by modulating the glottal air stream within the vocal tract (Fant, 1971; Stevens & House, 1955, 1961). For oral phonemes, the vocal tract may simply be viewed as a tube consisting of the pharyngeal and oral cavities, and augmented for the production of nasal phonemes by an additional branched tube coupled to the pharyngeal and oral cavities. The ability to control coupling of the nasal cavities to the pharyngeal and oral cavities is crucial for the production of normal speech: Inability to decouple the nasal cavities from the remainder of the vocal tract will result in severely distorted speech. In addition, speakers must be able to control with some precision the timing of alternating these coupled and decoupled configurations of the vocal tract, to realize phonemic distinctions between nasal and oral segments.

This chapter offers a description of the control system that governs the coupling and decoupling of these resonating cavities, beginning with a brief summary of the mechanisms for closing and opening the velopharyngeal port in speech, and then considering, in some detail, the effects of phonetic content on velar position. Following this phonetic-content description is a phonetic-context description of velar function, which is concerned with considering the interaction between velar movement patterns for proximate phonetic segments.

Phonetic context effects are interesting because of the insights they may provide into the form of the motor plan for speech: In what units is the motor program specified, and over what number of these units is it prepared? One way we may gauge the degree to which we understand a system (for example, the form of the motor plan employed for speech) is to build a model embodying the known facts, and then to examine the model's ability to predict the behavior of the natural system under novel conditions. The success of the model in predicting the behavior of the system is, then, an index of the caliber of our understanding. This is a time-honored test of great usefulness

---

*Chapter to appear in Speech and Language: Advances in Basic Research and Practice, Vol. IV, Norman J. Lass, Editor.

Acknowledgment. The experiment reported here, and the preparation of this manuscript, were supported by NINCDS Grants NS-13617 and NS-05332 and BRS Grant RR-05596. The research was carried out at the Haskins Laboratories, whose staff and facilities made this work possible. I want to thank Katherine S. Harris and Patrick W. Nye for their thoughtful reading of the manuscript and very helpful suggestions. Finally, I especially want to thank Thomas Baer for his invaluable and tireless comments during the preparation of this manuscript.

and therefore, employing the velar coarticulation data reported in the literature, as well as data from an experiment to be reported here, we propose to offer a model of velar function that may prove to be a useful subject for further comparisons with the actions of the human articulatory system.

II. MECHANISMS OF VELAR CONTROL

A. Introduction

The role of the velopharyngeal mechanism in speech has been of interest for many years, but the history of this interest will only be surveyed briefly in this chapter. (See Dickson & Maue-Dickson, 1980, for a comprehensive historical perspective.) Thus, Fritzell (1969) reports studies by Czermak (1857, 1858, 1869) and Passavant (1863) involving both indirect and direct measures of velopharyngeal closure during speech.[1] The conclusion of these experiments was that velar height decreases through the vowel series [i], [u], [o], [e], [a]. Passavant also placed tubes of varying diameters in the velopharyngeal port region to determine how small the port must be to prevent nasalization of oral speech sounds, and found that a cross-sectional area of $12.6 mm^2$ had little effect on speech quality, but that a cross-sectional area of $28.3 mm^2$ resulted in the nasalization of most consonants. He also reported a bulging in the posterior pharyngeal wall, above the level of velopharyngeal closure, during the speech of a cleft palate speaker. He assumed that this bulging, which has come to be known as Passavant's ridge, occurs in all speakers.

It is possible to trace two lines of investigation leading from these early studies. The first line concerns the dimensions and mechanisms of oral and nasal articulation. More specifically, is oral articulation achieved by: (a) posteriorly and superiorly directed movement of the velum; (b) a combination of velar movement and anteriorly directed movement of the posterior pharyngeal wall (Passavant's ridge); or (c) a combination of velar movement and medially directed movement of the lateral pharyngeal wall? Which muscles are responsible for closing the velar port? Need the port be completely closed for all "oral" articulations? And, is nasal articulation achieved by the contraction of some muscle or muscle group, or solely by decreasing activity in those muscles responsible for oral articulation? The second line of investigation concerns the nature of variations in velopharyngeal activity both as a function of the identity of phonetic segments and as a function of interactions among proximate segments (coarticulation).

B. Velopharyngeal Closure Mechanisms

It is generally accepted that the levator palatini is the muscle responsible for elevating and retracting the velum (cf. Bell-Berti, 1976; Bosma, 1953; Dickson, 1975; Fritzell, 1969; Lubker, 1968). This upward and backward motion of the velum is observed in all normal speakers.

The questions concerning velopharyngeal closure mechanisms that continue to receive attention and will briefly be considered here involve the roles of the posterior and lateral pharyngeal walls in the closing gesture. The first of these, the question of the existence and ubiquity of Passavant's ridge as a mechanism for closing the velar port, has been addressed by a number of people. For example, Calnan (1957) has disputed the presence of Passavant's ridge in most speakers and claimed that such a mechanism would be far too sluggish and fatigable to be a reliable compensatory mechanism for speakers with inadequate palatal musculature. Hagerty and colleagues (Hagerty, Hill, Pettit, & Kane, 1958; Hagerty & Hill, 1960) concluded that Passavant's ridge is not a mechanism used by most normal speakers, although post-operative cleft palate subjects tend to use more posterior pharyngeal wall movement in speaking than do normal subjects. Carpenter and Morris (1968) concluded that, when Passavant's ridge occurs in speakers with surgically repaired clefts, it may be used as a reliable compensatory mechanism for some of them. In parallel studies of normal and cleft palate speakers, Björk (1961) and Nylen (1961), respectively, found that normal speakers did not use anteriorly directed movement of the posterior pharyngeal wall in closing the velar port, and that among cleft palate speakers judged to have no insufficiency, velar movement patterns were comparable to those of normal speakers. A Passavant's ridge was identified in 11 of Nylen's 27 speakers whose velopharyngeal closure was judged to be inadequate for speech.

Observations of anteriorly directed movements of the posterior pharyngeal wall have been attributed to contraction of the superior pharyngeal constrictor. Similarly, the regularly observed medial movements of the lateral pharyngeal walls, at the level of velopharyngeal closure, have also been attributed to the action of this muscle (cf. Fritzell, 1969; Lubker, 1968; Shprintzen, Lencione, McCall, & Skolnick, 1974; Skolnick, McCall, & Barnes, 1973; Zagzebski, 1975). However, this view is difficult to support anatomically because the superior margin of that muscle is at or below the palatal plane (Dickson, 1975), and velopharyngeal closure is frequently above this level. It therefore seems unlikely that the superior pharyngeal constrictor can be responsible for these movements. Furthermore, the converging movements of the lateral walls and velum are strikingly parallel in both time course and extent (cf. Harrington, 1944; Niimi, Bell-Berti, & Harris, 1978; Skolnick, 1969; Zagzebski, 1975). Finally, the weight of evidence from electromyographic studies on the role of the superior pharyngeal constrictor in closing the velar port is divided, with supportive data reported by Fritzell (1969) and Lubker (1968) and conflicting data reported by Bell-Berti (1973, 1976) and Minifie, Abbs, Tarlow, and Kwaterski (1974).

An alternative view is that both lateral pharyngeal wall movement and velar elevation and retraction are caused by contraction of the levator palatini (cf. Bell-Berti, 1973, 1976; Bosma, 1953; Dickson, 1975; Dickson & Dickson, 1972; Honjo, Harada, & Kumazawa, 1976; Niimi et al., 1978). However, some investigators (cf. Shprintzen et al., 1974; Skolnick et al., 1973) have claimed that because the localized bulge in the lateral walls occurs below the level of the "levator eminence" (on the superior surface of the velum), the bulge cannot result from contraction of the levator palatini. The studies of Azzam and Kuehn (1977) and of Dickson (1972), though, indicate that the "levator eminence" may result from contraction of the uvular muscle, and not of the levator palatini, thus casting doubt on the validity of the argument.

It is not clear, then, whether or not the superior pharyngeal constrictor plays a role in closing the velar port for speech. It does, however, seem reasonable to attribute to it, and to the middle pharyngeal constrictor as well, some portion of the lateral pharyngeal wall movement observed in the oropharynx for open vowels (cf. Minifie, Hixon, Kelsey, & Woodhouse, 1970; Zagzebski, 1975). This seems especially reasonable in light of EMG data showing parallel activity in the pharyngeal constrictor muscles, at the level of the epiglottis and at the superior boundary of the superior pharyngeal constrictor, for speech (Bell-Berti, 1973, 1976).

C. Velopharyngeal Closure: Critical Port Size

A second question raised by studies of velar port control is whether the port must be completely closed for all oral phonemes, to prevent coupling of the nasal and oral cavities. In experiments with synthesized speech, House and Stevens (1956) varied the ratio of the driving point impedance of the velopharyngeal port (which is a function of the port's cross-sectional area) to the internal impedance of the vocal tract, and found that nasal coupling increased as this ratio decreased. They reported that listeners failed to judge any of their vowel stimuli produced with a port area of $25mm^2$ as "more nasal" than those produced with the port completely closed, but that high vowels produced with a port area of $71mm^2$ (the next larger area in their series) were judged as "more nasal" than those produced with the smaller area.

Björk's (1961) report provides us with a useful rule-of-thumb for estimating port area from lateral view x-ray pictures. He found the cross-sectional area of the port to be a linear function of the port's sagittal minor axis, and that the area may be computed by multiplying the antero-posterior dimension of the port (expressed in mm) by 10. Applying Björk's computation to antero-posterior dimension data available in the literature, we find, in general, that speakers having minimum velar port areas of less than about $30mm^2$ had speech that was nearly normal, while those having greater minimum port areas had speech judged as being nasalized. Indeed, the larger the minimum port area, the more seriously distorted was the speech (cf. Nylen, 1961; Subtelny, Koepp-Baker, & Subtelny, 1961). In agreement with these data are those of Warren's (1967) study of nasal air flow as an estimate of velar port size: speech was judged adequate at minimum port areas under $20mm^2$, and inadequate when the minimum port area was greater than $20mm^2$. In agreement with the results of the speech synthesis and physiological studies are the results of Isshiki, Honjow, and Morimoto (1968), who induced velopharyngeal incompetence in their subjects by placing polyvinyl tubes in their velar ports, and found the critical port area to be about $20mm^2$.[2]

Thus, complete closure of the port is not always required for normal speech production. The speaker need only make the port sufficiently small so as to establish admittances into the nasal, oral, and pharyngeal branches, at the velar port, that will prevent the nasal branch from affecting the overall vocal tract transfer function for sonorants. For obstruents, the port must also be sufficiently small to prevent nasal air flow. Indeed, Björk reports the presence of a gap between the velum and posterior pharyngeal wall during the production of some obstruent segments judged as completely normal. (See the Appendix for a discussion of the acoustical theory of nasality.)

D. Velopharyngeal Port Opening Mechanisms

A third question is how the velar port is opened to permit nasal coupling. There are two ways in which the velar port could be opened. The first, and simplest, is that the muscles used in closing it relax and the elastic tissue forces open the port. The second possibility is that the contraction of some muscle or group of muscles (possibly palatopharyngeus or palatoglossus) pulls downward on the velum while the muscles involved in closing the port are relaxing.

In an EMG study, Fritzell (1969) found palatopharyngeus activity to vary across subjects, but in general to be more active for the vowel [a] than for [i] and [u]. Bell-Berti (1973, 1976) has reported that the palatopharyngeus works synergistically with the levator palatini, but that it is more active for open than for close vowels, apparently acting to narrow the faucial isthmus for these articulations. Thus, the available EMG data do not provide support for the role of palatopharyngeus as a velar depressor.

The situation is less transparent, however, for the palatoglossus. Several studies have reported that palatoglossus activity occurs when levator palatini activity is suppressed; that is, at times corresponding to nasal consonant articulation (cf. Benguerel, Hirose, Sawashima, & Ushijima, 1977; Fritzell, 1969; Lubker, Fritzell, & Lindqvist, 1970; Lubker, Lindqvist, & Fritzell, Note 2). In contrast, however, Bell-Berti (1973, 1976; Bell-Berti & Hirose, 1973) has reported EMG data, recorded from several speakers, showing no difference in palatoglossus activity associated with changes in the status of the velar port. Instead, these data show palatoglossus activity for high back vowels and velar consonants, speech segments for which levator palatini activity is also high (see also Kuenzel, 1978), indicating palatoglossus involvement in tongue-dorsum elevation. These authors have also reported recording palatoglossus activity for low vowels, presumably to narrow the faucial isthmus. Finally, Bell-Berti and Hirose (1973) have reported data for one speaker who apparently uses the palatoglossus in both tongue-dorsum elevation and velum-lowering gestures.

Taken together, these data suggest, at the least, that there is no universal mechanism for lowering the velum involving increased activity in any muscle (Bell-Berti, 1976). Rather, the basic mechanism for opening the velar port involves the suppression of activity in those muscles acting to close it, and for some speakers the contraction of the palatoglossus to provide a supplementary downward force. There is no evidence that the palatopharyngeus ever provides such a force.

III. THE EFFECTS OF PHONETIC CONTENT

Closely related to the question of how the velopharyngeal port is closed to achieve oral articulation is the question of how tightly closed it must be, for a given segment type, to prevent nasal coupling. This question is obviously related to the effect of phonetic content upon velar height. However, these two aspects of the question will be considered separately, to insure a thorough appreciation of the segmental effects.

Moll (1962), and others, have concluded that velar port closure and, hence, velar elevation, are greater for high vowels than for low vowels and that closure is incomplete for vowels in nasal environments. One explanation given for these differences in articulator position includes the mechanical constraints within the articulatory system and changes in the timing relationships among the control signals to the articulators (cf. Lindblom, 1963; Stevens & House, 1963). Thus, one possible description of velar position control might be an 'on-off' algorithm, with variable control-signal timing relationships and a correction for mechanical constraints.

However, this view has been disputed by the evidence of a number of studies (cf. Bell-Berti, 1976; Fritzell, 1969; Lubker, 1968; Moll & Shriner, 1967). For example, Fritzell (1969) and Lubker (1968) reported a high correlation between velar position and velar EMG activity for vowels of different height, with greater elevation and EMG potentials for high vowels than for low vowels. These data, and others not enumerated here, confirm the reports of Czermak (1857, 1858, 1869) and of Passavant (1863), that palatal height increases through the series [a], [e], [o], [u], [i].

Extending our view to consonantal segments, we find, not surprisingly, that nasal consonants have the lowest velar position and smallest levator palatini EMG potentials of any speech sounds (cf. Bell-Berti, 1976; Bell-Berti, Baer, Harris, & Niimi, 1979; Fritzell, 1969; Lubker, 1968). Conversely, obstruent consonants have the highest velar elevation and largest levator palatini EMG potentials (cf. Bell-Berti, 1976; Bell-Berti & Hirose, 1975; Harris, Schvey, & Lysaught, 1962; Lubker et al., 1970).

It is clear from the data of many studies, carried out over more than a century on several different languages, that it is possible to make at least one general statement about the relationship between velar elevation and the phonetic content of a piece of speech: Velar elevation and levator palatini EMG potentials for oral speech sounds vary directly with the degree of oral cavity constriction, decreasing through the series: obstruents--close vowels--open vowels. In addition, the results of perceptual tests of the effects of opening the velar port reveal that oral consonants are distorted at smaller port areas than are close vowels, which in their turn are perceived as being "nasal" at smaller port areas than are open vowels. Since velar elevation decreases through this same series, we might conclude that speakers recognize the acoustic consequences of inappropriately large velar port areas and modify velar port area (by controlling velar elevation) to avoid introducing the distortions of nasal coupling.

However, some disagreement remains about levator palatini EMG-potential relationships and velar position relationships within the group of obstruent consonants. It has been suggested that those consonants characterized by high intraoral air pressure levels (e.g., the high intensity voiceless fricatives) are produced with the strongest levator palatini EMG potentials (cf. Lubker et al., 1970). There are, however, reports of velar function differences among speakers, differences indicating that the voiceless obstruents are produced with the strongest levator palatini activity only by some speakers (Bell-Berti, 1973, 1975; Bell-Berti & Hirose, 1975). These differences among speakers are systematic, and are related to the different articulatory strategies used by the speakers to maintain voicing during obstruent consonant production (cf. Bell-Berti, 1975). Thus, some speakers regularly use greater levator palatini activity (and, consequently, higher velar elevation) for voiced obstruents than for their voiceless cognates, increasing the volume of, and decreasing the supraglottal pressure in, the pharyngeal cavity. This adjustment maintains the transglottal pressure difference required for glottal pulsing to continue during the period of vocal tract occlusion for obstruent production (cf. Bell-Berti, 1975; Perkell, 1969; van den Berg, 1958). Conversely, some speakers maintain the transglottal pressure difference necessary for glottal pulsing by allowing air to 'leak' through a partially opened port (Dixit & MacNeilage, Note 1). Still other speakers accomplish this vocal tract adjustment in other ways, including advancing and depressing the tongue root, depressing the larynx, or increasing oral cavity volume (cf. Bell-Berti, 1975; Fujimura, Tatsumi, & Kagaya, 1973; Kent & Moll, 1969).

These secondary articulatory maneuvers controlling effective pharynx volume, as well as the adjustment of pharyngeal cavity cross-sectional area for vowels (cf. Bell-Berti, 1973), are important for two reasons. First, and most obvious, is that an adequate model of speech production must account for all of the articulatory activities of the speech mechanism. Second, and perhaps of more direct relevance here, their interaction with port-closing gestures might otherwise confuse our interpretation of data collected during the production of long sequences of segments, which we must collect if we are to improve our understanding of the interaction between motor plans for, and/or the execution of, speech.

IV. THE EFFECTS OF PHONETIC CONTEXT

In addition to describing the mechanisms of oral and nasal articulation and their interaction with phonetic content, studies of velar function have also tried to define, usually in terms of segmental units, the extent of the influence of velar position for one segment on velar position for proximate segments, to gain insight into the size of the units of the speech motor plan. Most often, the focus has been on the influence of velar position for nasal consonants on velar position for vowels. Indeed, it is a common observation that vowels adjacent to nasal consonants are nasalized (cf. Leutennegger, 1963, p. 150), and, more specifically, that nasality is assimilated in vowels before nasal consonants (Bronstein, 1961, p. 109). Ohala (1971) has reported greater nasal coarticulation effects in vowels before than in vowels following nasals, and states that velar lowering begins as soon as elevation is no longer required for obstruent articulation. Ushijima and Sawashima (1972) found that vowels in nasal environments have lower velar positions than do the same vowels in oral environments, and that the greatest velar elevation occurs for obstruent consonants immediately following nasals. In a study having a somewhat different objective, one describing the effects of vowel environment on velar position for consonants, the velum was found to be higher for both oral and nasal consonants ocurring in close-vowel, than in open-vowel, environments (Bell-Berti et al., 1979).

In an account of a study of the timing of velar movements in relation to other, segmentally defined, articulator movements, Moll and Daniloff (1971) reported that movement toward opening of the velar port began during articulator movement toward the first vowel in CVN and CVVN sequences. In NC and NCN sequences, movement toward closure began during the first nasal consonant. In NVC sequences, movement toward closure was quite similar to that for NC sequences, although it began a bit later in the former and closure was not always complete during the vowel.

One general model of speech production that has been tested with velar function data is Henke's (1966) phoneme-based model. This model assumes the input to the articulatory system to be a string of phonemes that are specified as sets of invariant articulatory goals, or "features." It postulates a "look-ahead" procedure that allows the goals of phonemes occurring later in the string to influence the current and intervening vocal tract configurations, so long as these anticipated goals are not in conflict with any more immediate goals.[3] A model developed from the Moll and Daniloff data proposes two velar port goals: 'closed' for oral consonants and 'open' for nasal consonants. In this scheme, velar position for vowels is assumed to be unspecified, and determined by the next specified position. The predictions of this essentially binary model agree with those of Henke's model of speech production, and a substantial proportion of the data are in agreement with the predictions of such a look-ahead model.

There are, however, at least three instances in which blind application of the look-ahead model fails to account for observations of human speech. The first of these is the reported effect of a marked junctural boundary in blocking anticipation of a downstream goal (McClean, 1973; Ushijima & Hirose, 1974). McClean suggests that the delay in nasal anticipation may result from a high-level reorganization of commands to the velum, and that this explanation is consistent with a look-ahead model.

The second discrepancy between the data and the look-ahead model concerns predictions of timing. For example, in NC sequences, velar movement toward closure often begins before the oral constriction for the nasal consonant is achieved. Kent, Carney, and Severeid (1974) suggest that the binary model need only be modified to allow a motor program that simultaneously issues commands to different articulators for different segments.

The third, and to this view the most serious, failure of the binary model concerns the prediction of velar height for vowels in utterances whose consonants are either all oral or all nasal. In such phoneme sequences velar height is not constant, as the model predicts, but rather decreases for vowels occurring within oral consonant environments (Bell-Berti, 1979) and increases for vowels occurring within nasal consonant environments (Kent et al., 1974), in direct contradiction with the prediction that the velar goal for the consonants will be anticipated during the vowels.

Finally, there are two additional problems surrounding the development of an adequate model of velar function that stem from limitations in the quality of many of the existing data. These limitations in their turn result from shortcomings in the design of many of the experiments. The first of these is that the restricted nature of the phonetic inventory in the speech samples that have been studied renders impossible many of the comparisons between oral- and nasal-environment effects that might reveal the segmental, or temporal, extent of the coarticulatory field. That is, since it has been assumed that nasality is the only phonetic feature whose presence will influence velar height for non-nasal segments, nearly all of the speech samples contain nasal segments. Those sequences not containing nasal segments are contrasted with utterances that do contain nasals, and not with other, minimally contrastive, non-nasal utterances.

A second, and more serious, limitation is imposed by the tacit assumption that velar position for vowels between oral consonants will be the same as velar position for the oral consonants, in face of the substantial body of contrary data indicating that velar position for oral speech sounds varies directly with the oral cavity constriction for those sounds (cf. Bell-Berti, 1973, 1976; Czermak, 1857, 1858, 1869; Fritzell, 1969; Lubker, 1968; Moll, 1962; Passavant, 1863). That this assumption has often been made is evident in the criteria for establishing the beginning of anticipatory influences of nasal consonants on preceding vowels, usually taken as the earliest observation of velar lowering after peak elevation for the oral consonants in CVN sequences. It is obvious, however, from the data of Figure 1 that the velum lowers for vowels following obstruent consonants even when those vowels occur in entirely oral environments. Thus, it is impossible to estimate the extent of the anticipatory field from measures of the earliest moment of velar lowering in CVN sequences, since this lowering may be associated with the velar-position specification for the vowel. Rather, descriptions of the timing of anticipatory nasal coarticulation must derive from comparisons of velar position for vowels in both oral and nasal environments.

V. A SPATIAL-TEMPORAL MODEL OF VELAR FUNCTION

A. Preliminaries

The model offered here is intended to account for observations of velar position and the timing of velar movements in normal speech. This model assumes that the levator palatini is the muscle primarily responsible for velopharyngeal closure and that the strength of levator palatini contraction is reflected fairly directly in velar position. This assumption is based on the knowledge that the area of the velopharyngeal port is closely related to the position of the velum, with port area decreasing directly with increasing velar elevation (Ushijima & Sawashima, 1972). In addition, we know the levator palatini muscle to be responsible for raising and retracting the velum in the port-closing gesture (cf. Bell-Berti, 1976; Fritzell, 1969; Lubker, 1968). However, since upward movement of the velum may continue above the level at which the port closes completely, measures of velar elevation more directly reflect the motor commands underlying velar gestures than do measures of velar port area.

The data on which this model rests include electromyographic and positional information recorded from the velum, much of which has been reported elsewhere (cf. Bell-Berti, 1973, 1976, 1979; Bell-Berti et al., 1979; Bell-Berti & Hirose, 1975). Briefly, EMG recordings from the levator palatini have shown the magnitude of its EMG potentials to correlate highly with changes in velar position (Bell-Berti & Hirose, 1975), within a constant phonetic environment. These potentials are greatest for obstruent consonants, smaller for close vowels, smaller still for open vowels, and lowest for nasal consonants (cf. Bell-Berti, 1973, 1976). Velar height decreases through the same series, highest for obstruents and lowest for nasals (cf. Bell-Berti et al., 1979; Bell-Berti & Hirose, 1975).

In addition, velar position data were collected in an experiment to supplement existing data, providing information on coarticulation within entirely oral utterances. These data permit one to examine the temporal

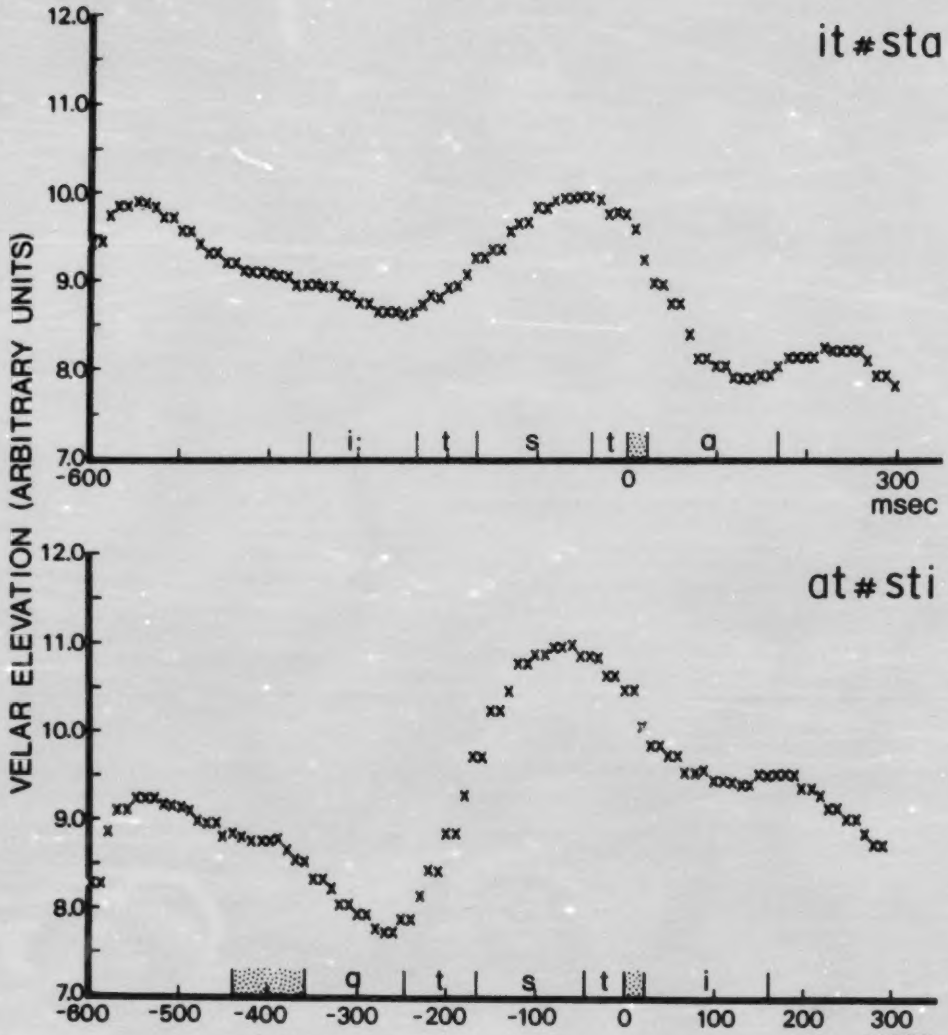


Figure 1. Ensemble-average velar elevation functions for two $V_1C_nV_2$ phrases from the utterance set described in Section V,B,1, spoken in the carrier sentence "It's a ________ again." The upper figure contains the function for the phrase [flit#stap]; the lower figure contains the function for the phrase [kat#stiz]. Velar elevation is given in arbitrary units, time in msec. Average duration of the segments of $V_1t\#stV_2$ are displayed beneath each function. Zero on the abscissa represents the acoustic end of the consonant string.

extent of interaction effects among vowels and consonants, in entirely oral utterances, and are described below.

B. The Experiment

1. Method

The subject in this study was a native speaker of standard Greater Metropolitan New York City English. The experimental utterances were 27 two-word phrases having the general form $V_1C_nV_2$. V_1 and V_2 were [i] and [a], respectively, in 15 of the phrases, and the reverse in the remaining 12 phrases. C_n consisted of combinations of [s] and [t], with word-boundary positions systematically varied in each of the vowel-order sets. This produced such contrasts as, for example, [it#sta], [at#sti] and [ats#ti]. Nine minimal contrasts were possible between vowel-order sets, in addition to the possible contrasts within each vowel-order set among utterances having consonant strings of different duration (and number of segments). Each phrase began and ended with an obstruent consonant, although different consonants began and ended the two sets. The 27 phrases were embedded in the carrier sentence "It's a __________ again," and placed in lists in random order. The lists were repeated until the subject had produced from five to eight tokens of each.

A flexible fiberoptic endoscope (Olympus VF Type O) was inserted into the subject's nostril, and positioned so that it rested on the floor of the nasal cavity with its objective lens at the posterior border of the hard palate, providing a view of the velum and lateral nasopharyngeal walls, from the level of the hard palate to the maximum elevation of the velum. A long thin plastic strip with grid markings was also inserted into the subject's nostril and placed along the floor of the nasal cavity and over the nasal surface of the velum, to enhance the contrast between the edge of the supravelar surface and the posterior pharyngeal wall.

Motion pictures of the velum were taken through the endoscope at 60 frames per second. The position of the high point of the velum was then tracked, frame-by-frame, with the aid of a small laboratory computer. The measurements of velar elevation for the tokens of each utterance type were aligned with reference to the acoustic boundary between the end of the consonant string and beginning of the second vowel, and frame-by-frame ensemble averages were calculated. Vowels and medial consonant durations were measured from the digitized audio waveforms of the speech samples of each repetition.

2. The Data

First, there are two general, qualitative observations that can be made about these data. The first, and most striking, is that the velum continues to rise throughout consonant strings of considerable length--as many as 5 segments and as long as 360 msec--occurring in oral environments. This characteristic of velar behavior illustrates both the nature of the speech motor program and the size of the motor program units, and suggests that articulatory gestures may be programmed as movements and not as fixed articulatory targets or goals. Alternatively, the individually-specified positional goals for segments may sum cumulatively, and even the most extreme goal may be exceeded. Yet another alternative, again one assuming positional goals, is that the velar goal may not be achieved even during the production of a string of five obstruent segments having a duration of 360 msec. Implicit in this last hypothesis is a velar position goal that far exceeds the velar position necessary to prevent nasal coupling.

The second observation, already mentioned briefly above and which admittedly cannot be separated from the first, is that velar postion for vowels differs from velar position for oral consonants. The obvious conclusion, therefore, is that the velar goals for vowels differ from those for consonants. Furthermore, the goals for open and close vowels, at the least, may very well differ from each other.

Several more specific, quantitative observations are also possible. One observation concerns differences in velar position for different vowels. Another concerns the relationships between vowel environment and maximum velar elevation for a consonant string. Still other observations are concerned with the time course of velar elevation and lowering in relation to other articulatory, and acoustic, events.

Turning attention first to velar position for the vowels [i] and [a], elevation was greater for [i] than for [a] in each of the 18 possible (nine first- and nine second-syllable) comparisons ($t_{17}=2.30$, $p<.05$). These differences, seen in Figure 2, were more pronounced in the second than in the first syllable (V_2: $t_8=4.95$, $p<.01$; V_1: $t_8=1.88$, $p>.05$), possibly reflecting differences between syllables in lexical stress and/or the phrase-initial or phrase-final consonant.

Vowel environment had a significant influence on velar elevation for consonants: Peak elevation was greater for [aC_ni] than for [iC_na] phrases in all minimal comparisons, and on average (12 [aC_ni] and 15 [iC_na] phrases). The average difference in peak elevation, between vowel-order sets, was highly significant ($t_{25}=6.24$, $p<.001$), and indicates that the influence of V_2 on peak elevation for consonants is greater than that of V_1 (Figure 3). Since the peak in the velar elevation function is nearer to V_2 than V_1, this difference in vowel influence may simply reflect the temporal proximity of the beginning of V_2 to the velar elevation peak. On average, peak elevation occurs 75 msec before the (acoustic) beginning of the second vowel, and the average duration of the medial consonant strings is 226 msec.

In addition to being conditioned by the following vowel, peak velar elevation is also strongly influenced by the duration of the medial consonant string, within each vowel-order set (Figure 4). Thus, there is a strong positive correlation between the duration of the consonant string and maximum elevation, with $r=.74$ for the [aC_ni] phrases and $r=.86$ for the [iC_na] phrases. The lower correlation for the former probably relects the smaller range of peak velar elevations within that group. This reduced range may, in turn, be the result of mechanical constraints that impose ceiling effects on velar elevation possibilities. That is, velar elevation was already so extreme that even large increases in levator palatini contraction could not produce substantial increases in elevation.

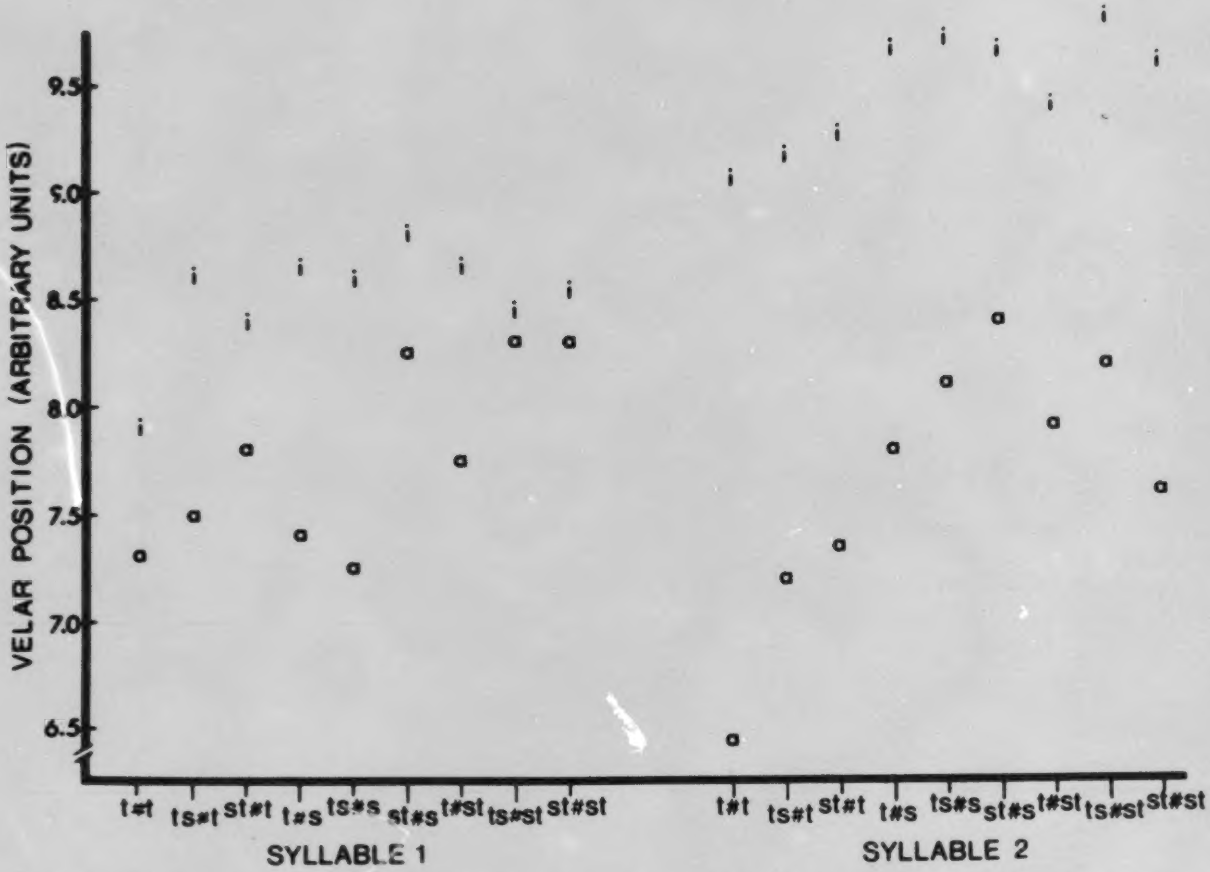


Figure 2. Velar position minima for the vocalic portions of the first and second syllables of the phrases described in Section V,B,1. Velar elevation is given along the ordinate in arbitrary units. Minimal-contrast phrases are represented along the abscissa by their consonant strings; syllable 1 is at the left, syllable 2 is at the right.

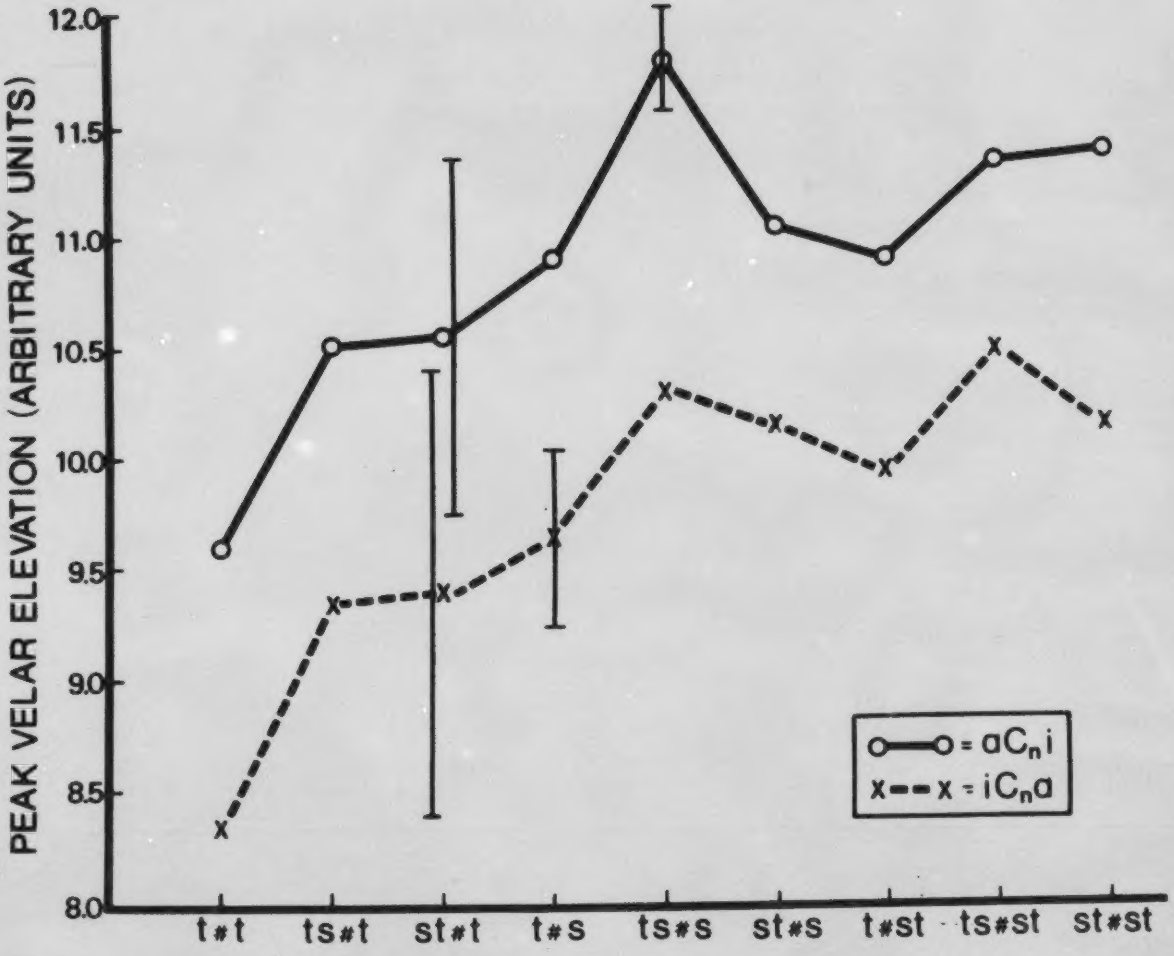


Figure 3. Peak velar elevation, from the ensemble averages, for minimal contrasts indicated along the abscissa. The smallest and largest standard deviation values are shown bracketing their respective means.

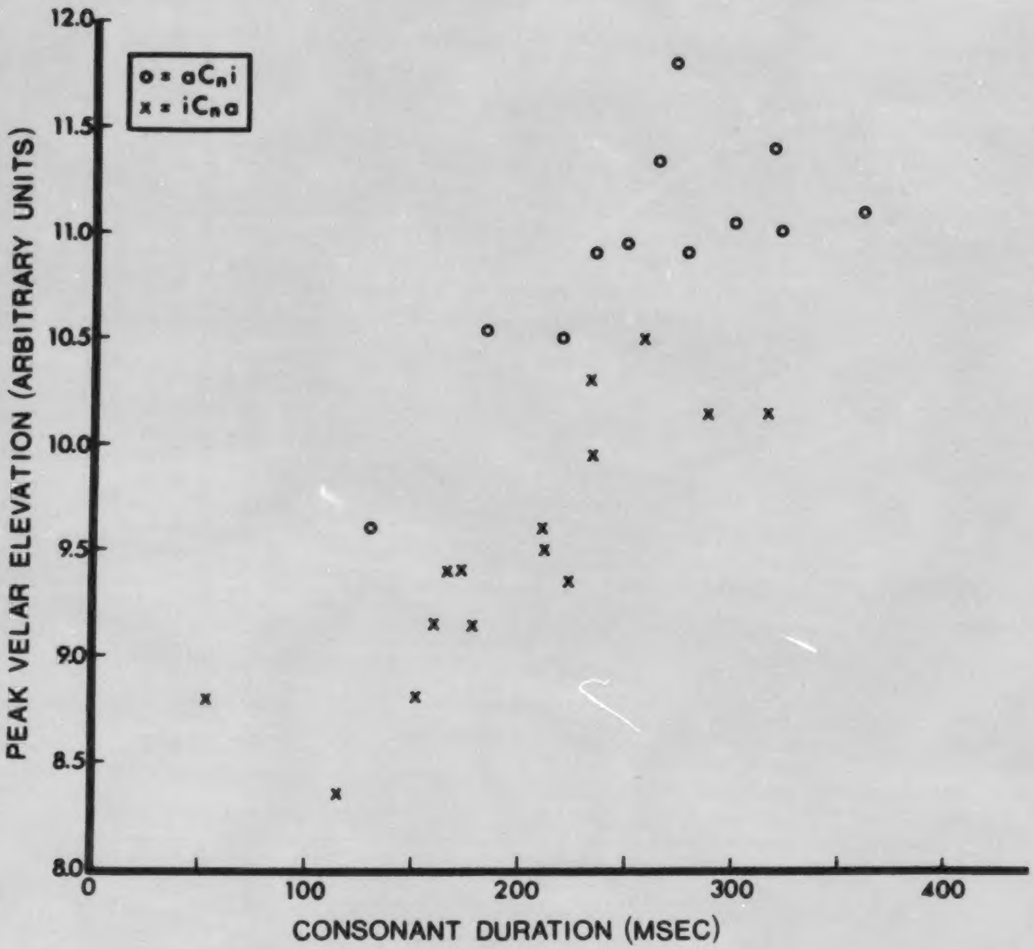


Figure 4. Scatter-plot of peak velar elevation (along the ordinate) vs. consonant-string duration (along the abscissa).

Finally, to estimate the time at which V_2 exerts more influence on peak elevation than does V_1, velar elevation was compared in the nine minimal pairs at several times before peak elevation was achieved: at 100, 150, 200, and 250 msec before the beginning of V_2. The mean difference in velar position was determined for each time point by subtracting the value obtained for $[iC_na]$ strings from that obtained for $[aC_ni]$ strings. So long as V_2 exerts the greater influence, this difference should be positive, and it should decrease as the influence of V_2 diminishes, becoming negative when the influence of V_1 exceeds that of V_2. These data are summarized in Table 1. Clearly, at even 100 msec before V_2, the influence of that vowel is small ($t_8=1.19$), and at 200 msec before V_2 the mean difference across comparison pairs is negative, indicating that the influence of V_1 predominates.

Table 1

Mean difference in velar position between $/aC_ni/$ and $/iC_na/$ utterances, taken at 50 msec intervals before the (acoustic) end of the consonant string ($t=0$ msec). The difference is greatest at $t=50$ msec, where V_2 exerts the greater influence, and smallest at $t=250$ msec, where the influence of V_1 is greater.

comparison time (msec before V_2)	(50)	100	150	200	250
Mean Difference	116.3	78.8	45.7	-11.7	-62.7
$t_8=$	5.08	1.19	.78	.20	1.15
p<	.001	.1	.1	.1	.1

C. The Model

This n-ary model of velar function postulates the segment-by-segment specification of both spatial and temporal parameters, permitting the description both of the data presented here and of those already in the literature, and generating hypotheses readily open to evaluation.[4] This model requires the specification of at least four positional or movement goals, one each for nasal consonants, open vowels, close vowels, and obstruent consonants. Additional spatial goals may be required for half-close vowels and sonorant consonants; it should, on the other hand, be possible to specify velar position for nasal vowels as an interaction between the nasal consonant and the appropriate close or open vowel goals. Thus, velar position for nasalized close vowels is expected to be higher than that for nasalized open vowels.

The remaining differences in velar position, those resulting from coarticulatory interactions, would be accounted for with the temporal parameter, with the model positing that each successive velar goal is initiated some fixed time before the (acoustic onset of the) segment for which it is specified. The velar gesture is also assumed to end gradually, rather than abruptly, and to be completed some fixed time after the (acoustic) end of the segment for which it is specified. The model assumes that the velum is programmed to achieve its maximum excursion for a segment before the (acoustic) end of the segment. Once the velum has achieved this maximum displacement, it moves either towards its rest position or, possibly, some neutral, speech-ready position (cf. Chomsky & Halle, 1968, p. 300). (It should be possible to determine whether or not this movement away from the maximum displacement is toward the rest position or the 'neutral' position by comparing velar movement patterns just before marked junctural boundaries, where the neutral position might be expected, and in utterance-final positions, where the rest position would be expected.) The goal specification may take the form of movements toward and away from some spatial target position, or, alternatively, simply of movements of greater or less extent. The present model is not able to distinguish between these two alternatives. In either case, however, the edges, or "tails" of the successive goal specifications overlap, producing the coarticulatory effects commonly described.

The model predicts that the vowel following a consonant string will have greater influence on peak velar elevation than will a preceding vowel because the peak in the elevation function occurs late in the string; that is, closer to the second vowel (see Figure 1). This prediction is, indeed, supported by the data offered above, where peak elevation is greater in /aC$_n$i/ than in /iC$_n$a/ phrases. Similarly, velar position in the earlier portion of the consonant string is expected to be more heavily influenced by the first vowel, a prediction again supported by these data. Differences in velar position during nasal consonants would similarly be affected by the state values for adjacent segments, an hypothesis supported by the data of Bell-Berti et al. (1979).

The assumption of segments as the units of the motor program rests on several observations. First, the programmed unit is presumed to be no larger than a segment because velar elevation continues to increase through obstruent consonant strings of considerable length, and the peak elevation achieved is proportional to overall consonant duration. It seems unreasonable to assume that the velar goal for such strings is so much greater than would be necessary to prevent coupling that it is never reached. On the other hand, if one assumes a cumulative, segmental specification, this continuing elevation is to be expected.[5] Second, peak velar elevation occurs at a nearly constant time before the end of the consonant string; that is, it does not occur earlier in longer strings, as might be expected if the goal for the following vowel, which is lower than that of the consonant string, begins to exert its influence. Finally, the second vowel begins to exert its influence at a relatively fixed time before its acoustic beginning. Thus, the beginning of the velar gesture for the vowel is linked to the beginning of other components of the vowel gesture itself, and is not free to begin at different times in different phonetic sequences, as a feature-based model would predict. Instead, the beginning of the vowel gesture is expected to begin later in longer consonant strings (that is, later with reference to the beginning of the consonant string) than in shorter ones, and marked junctural boundaries would have the apparent effect of delaying 'anticipation' because the segment being anticipated begins later, and thus its influence begins later.

It is important to note that this description of anticipation implies that it is the result both of temporally fixed relationships among the component gestures comprising a particular phonetic segment and of temporal overlap, or co-occurrence, of gestures for successive segments. These data do not permit determination of the effect of changes in lexical stress and in speaking rate on the timing of the beginning and end of the velar gesture in relation to the acoustic onset of the segment for which they are specified. Nor does this model contain hypotheses about the precise temporal relationships among the component gestures of a single phonetic segment. Thus, while it claims that a vowel begins to influence an immediately preceding consonant string about 150 to 200 msec before the acoustic onset of the vowel, it makes no claims about when the velar gesture begins in relation to the beginning of tongue-body movements for the vowel gesture, except to state that this relationship is constant for any given pattern of lexical stress and speaking rate.

Obviously, a complete model of velar function is not yet available. However, after the values for the temporal and spatial parameters have been established, it should be possible to extend the model to account for suprasegmental influences on velar position. Once this has been done, the model may be used to predict velar position in a wide variety of utterances, to determine the model's validity.

REFERENCE NOTES

1. Dixit, P., & MacNeilage, P. F. *Coarticulation of nasality: Evidence from Hindi*. Paper presented at the 83rd Meeting of the Acoustical Society of America, 1972.
2. Lubker, J. F., Lindqvist, J., & Fritzell, B. *Some temporal characteristics of velopharyngeal muscle function*. Paper presented at the Phonetics Symposium, University of Essex Language Center, 1972.

REFERENCES

Abramson, A. S., Nye, P. W. Henderson, J., & Marshall, C. W. The perception of an oral-nasal continuum generated by articulatory synthesis. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-57*, 17-38.

Azzam, N. A., & Kuehn, D. P. The morphology of musculus uvulae. *Cleft Palate Journal*, 1977, *14*, 78-87.

Bell, C. G., Fujisaki, H., Heinz, J. M., Stevens, K. N., & House, A. S. Reduction of speech spectra by analysis-by-synthesis techniques. *Journal of the Acoustical Society of America*, 1961, *33*, 1725-1736.

Bell-Berti, F. The velopharyngeal mechanism: An electromyographic study. *Haskins Laboratories Status Report on Speech Research*, 1973, 1-159 (Supplement).

Bell-Berti, F. Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America*, 1975, *57*, 456-461.

Bell-Berti, F. An electromyographic study of velopharyngeal function in speech. *Journal of Speech and Hearing Research*, 1976, *19*, 225-240.

Bell-Berti, F. The extent of a vowel's influence: Evidence from a study of velar function. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th Meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979, 79-82.

Bell-Berti, F., Baer, T., Harris, K. S., & Niimi, S. Coarticulatory effects of vowel quality on velar elevation. *Phonetica*, 1979, *36*, 187-193.

Bell-Berti, F., & Hirose, H. Patterns of palatoglossus activity and their implications for speech organization. *Haskins Laboratories Status Report on Speech Research*, 1973, *SR-34*, 203-209.

Bell-Berti, F., & Hirose, H. Palatal activity in voicing distinctions: A simultaneous fiberoptic and electromyographic study. *Journal of Phonetics*, 1975, *3*, 69-74.

Benguerel, A.-P., Hirose, H., Sawashima, M., & Ushijima, T. Velar coarticulation in French: An electromyographic study. *Journal of Phonetics*, 1977, *5*, 159-167.

Björk, L. Velopharyngeal function in connected speech. *Acta Radiologica*, 1961, *Supplement 202*.

Bosma, J. F. A correlated study of the anatomy and motor activity of the upper pharynx by cadaver dissection and by animatic study of patients after maxillo-facial surgery. *Annals of Otology, Rhinology, and Laryngology*, 1953, *62*, 51-72.

Bronstein, A. J. *The pronunciation of American English*. New York: Appleton-Century-Crofts, 1961.

Calnan, J. S. Modern view on Passavant's ridge. *British Journal of Plastic Surgery*, 1957, *10*, 89-113.

Carpenter, M. A., & Morris, H. L. A preliminary study of Passavant's Pad. *Cleft Palate Journal*, 1968, *5*, 61-72.

Chomsky, N., & Halle, M. *The sound pattern of English*. New York: Harper & Row, 1968.

Czermak, J. N. Ueber das Verhalten des weichen Gaumens beim Hervorbringen der reinen Vocale. *Wiener Akad. Sitzungsberichte*, 1857, *27* (cited in Fritzell, 1969).

Czermak, J. N. Ueber reine und nasalierte Vocale. *Wiener Akad. Sitzungsberichte*, 1858, *28* (cited in Fritzell, 1969).

Czermak, J. N. Wesen und Bildung der Stimm- und Sprachlaute. *Czermak's gesammelte Schriften*, Vol. 2, Leipzig: Wilhelm Engelman, 1869 (cited in Fritzell, 1969).

Dickson, D. R. Normal and cleft palate anatomy. *Cleft Palate Journal*, 1972, *9*, 280-293.

Dickson, D. R. Anatomy of the normal velopharyngeal mechanism. *Clinics in Plastic Surgery*, 1975, *2*, 235-248.

Dickson, D. R., & Dickson, W. M. Velopharyngeal anatomy. *Journal of Speech and Hearing Research*, 1972, *15*, 372-381.

Dickson, D. R., & Maue-Dickson, W. Velopharyngeal structure and function: A model for biomechanical analysis. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice*, *Vol. III*, 1980.

Fant, C. G. M. *Acoustic theory of speech production* (2nd ed.). The Hague: Mouton, 1971.

Fritzell, B. The velopharyngeal muscles in speech: An electromyographic and cineradiographic study. *Acta Oto-laryngologica*, 1969, *Suppl. 250*.

Fujimura, O., & Lindqvist, J. Sweep-tone measurements of vocal-tract characteristics. *Journal of the Acoustical Society of America*, 1971, *49*, 541-558.

Fujimura, O., Tatsumi, I. F., & Kagaya, R. Computational processing of palatographic patterns. *Journal of Phonetics*, 1973, *1*, 47-54.

Hagerty, R. F., & Hill, M. J. Pharyngeal wall and palatal movement in postoperative cleft palates and normal palates. *Journal of Speech and Hearing Research*, 1960, *3*, 59-66.

Hagerty, R. F., Hill, M. J., Pettit, H. S., & Kane, J. J. Posterior pharyngeal wall movement in normals. *Journal of Speech and Hearing Research*, 1958, *1*, 203-210.

Harrington, R. A study of the mechanics of velopharyngeal closure. *Journal of Speech and Hearing Disorders*, 1944, *9*, 325-345.

Harris, K. S., Schvey, M. M., & Lysaught, G. F. Component gestures in the production of oral and nasal labial stops. *Journal of the Acoustical Society of America*, 1962, *34*, 743 (Abstract).

Henke, W. L. *Dynamic articulatory model of speech production using computer simulation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, 1966.

Honjo, I., Harada, H., & Kumazawa, T. Role of the levator palatini muscle in movement of the lateral pharyngeal wall. *Archives of Oto-Rhino-Laryngology*, 1976, *212*, 93-98.

House, A. S., & Stevens, K. N. Analog studies of the nasalization of vowels. *Journal of Speech and Hearing Disorders*, 1956, *21*, 218-232.

Isshiki, N., Honjow, I., & Morimoto, M. Effects of velopharyngeal incompetence upon speech. *Cleft Palate Journal*, 1968, *5*, 297-310.

Kent, R. D., Carney, P., & Severeid, L. R. Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research*, 1974, *17*, 470-488.

Kent, R. D., & Moll, K. L. Vocal-tract characteristics of the stop cognates. *Journal of the Acoustical Society of America*, 1969, *46*, 1549-1555.

Kozhevnikov, V. A., & Chistovich, L. A. *Rech', Artikulyatsiya, i vospriyatiye*, trans. as *Speech: Articulation and perception*. Washington, D.C.: Joint Publications Research Service, 1966, *30*, 543. (Originally published, 1965.)

Kuenzel, H. J. Reproducibility of electromyographic and velographic measurements of the velopharyngeal closure mechanism. *Journal of Phonetics*, 1978, *6*, 345-352.

Leutenegger, R. R. *The sounds of American English*. Chicago: Scott, Foresman & Company, 1963.

Lindblom, B. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 1963, *35*, 1773-1781.

Lubker, J. F. An electromyographic-cineradiographic investigation of velar function during normal speech production. *Cleft Palate Journal*, 1968, *5*, 1-18.

Lubker, J. F., Fritzell, B., & Lindqvist, J. Velopharyngeal function: An electromyographic study. *Quarterly Progress and Status Report* (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm), 1970, *STL-QPSR 4*, 9-20.

McClean, M. Forward coarticulation of velar movement at marked junctural boundaries. *Journal of Speech and Hearing Research*, 1973, *16*, 286-296.

Mermelstein, P. Calculations of the vocal-tract transfer function for speech synthesis applications. In *Proceedings of the Seventh International Congress on Acoustics*, Vol. 3. Budapest: Akademiai Kiado, 1971, 173-176.

Minifie, F. D., Abbs, J. H., Tarlow, A., & Kwaterski, M. EMG activity within the pharynx during speech production. *Journal of Speech and Hearing Research*, 1974, *17*, 497-504.

Minifie, F. D., Hixon, T. J., Kelsey, A. A., & Woodhouse, R. J. Lateral pharyngeal wall movement during speech production. *Journal of Speech and Hearing Research*, 1970, *13*, 584-594.

Moll, K. L. Velopharyngeal closure on vowels. *Journal of Speech and Hearing Research*, 1962, *5*, 30-77.

Moll, K. L., & Daniloff, R. G. Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America*, 1971, *50*, 678-684.

Moll, K. L., & Shriner, T. H. Preliminary investigation of a new concept of velar activity during speech. *Cleft Palate Journal*, 1967, *4*, 58-69.

Niimi, S., Bell-Berti, F., & Harris, K. S. Dynamic aspects of velopharyngeal closure. *Haskins Laboratories Status Report on Speech Research*, 1978, *SR-55/56*, 45-62.

Nylen, B. O. Cleft palate and speech. *Acta Radiologica*, 1961, *Supplement 203*.

Ohala, J. J. Monitoring soft palate movements in speech. *Project on Linguistic Analysis Reports* (Phonology Laboratory, Department of Linguistics, University of California, Berkeley), 1971, *13*, J01-J015.

Passavant, G. *Ueber die Verschliessung des Schlundes beim Sprechen*. Frankfurt a.M.: J. D. Sauerländer, 1863 (cited in Fritzell, 1969).

Perkell, J. S. *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge, Mass.: M.I.T. Press, 1969.

Rubin, P., Baer, T., & Mermelstein, P. An articulatory synthesizer for perceptual research. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-57*, 1-16.

Scully, C., & Shirt, M. A. Oral and nasal outputs for vowels in non-nasal contexts. *Proceeding of the Ninth International Congress of Phonetic Sciences*. Vol. 1. Copenhagen: Institute of Phonetics, University of Copenhagen, 1979, 212.

Shprintzen, R. J., Lencione, R. M., McCall, G. N., & Skolnick, M. L. A three dimensional cinefluorographic analysis of velopharyngeal closure during speech and nonspeech activities in normals. *Cleft Palate Journal*, 1974, *11*, 412-428.

Skolnick, M. L. Video velopharyngography in patients with nasal speech, with emphasis on lateral pharyngeal wall motion in velopharyngeal closure. *Radiology*, 1969, *93*, 747-755.

Skolnick, M. L., McCall, G. N., & Barnes, M. The sphincteric mechanism of velopharyngeal closure. *Cleft Palate Journal*, 1973, *10*, 286-305.

Stevens, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view*. New York: McGraw-Hill, 1972, 51-66.

Stevens, K. N., & House, A. S. Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America*, 1955, *27*, 484-493.

Stevens, K. N., & House, A. S. An acoustical theory of vowel production and some of its implications. *Journal of Speech and Hearing Research*, 1961, *4*, 303-320.

Stevens, K. N., & House, A. S. Perturbation of vowel articulation by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, 1963, *6*, 111-128.

Subtelny, J. D., Koepp-Baker, H., & Subtelny, J. D. Palatal function and cleft palate speech. *Journal of Speech and Hearing Disorders*, 1961, *26*, 213-224.

Ushijima, T., & Hirose, H. Electromyographic study of the velum during speech. *Journal of Phonetics*, 1974, *2*, 315-326.

Ushijima, T., & Sawashima, M. Fiberscopic observation of velar movements during speech. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1972, *6*, 25-38.

van den Berg, J. Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research*, 1958, *1*, 227-244.

Warren, D. Nasal emission of air and velopharyngeal function. *Cleft Palate Journal*, 1967, *4*, 148-156.

Zagzebski, J. A. Ultrasonic measurement of lateral pharyngeal wall motion at two levels in the vocal tract. *Journal of Speech and Hearing Research*, 1975, *18*, 308-318.

FOOTNOTES

[1] It is possible to observe articulator movements associated with speech gestures in two fundamentally different ways (cf. Bell-Berti, 1973). The first of these, *direct viewing*, involves measurement of articulator position, for example, measuring the elevation of the velum over time. Such techniques include visual observation (using posterior rhinoscopy or endoscopy) and cinematography, cineradiography, ultrasonic echo recording, and photoelectric recording of reflected light.

The second group of methods, *indirect viewing*, involves measurements of the cause or result of articulator position or displacement, implying but not specifying articulator movements, including electromyographic, air flow, acoustic, and transillumination recordings.

[2] It is of some interest to note that all of these fairly recent data provide general confirmation of Passavant's (1863) report that a velopharyngeal port cross-sectional area of 12.6 mm^2 had little effect on the quality of speech, while a cross-sectional area of 28 mm^2 resulted in nasal coupling for oral speech sounds, and thus, in distorted speech.

[3] Another frequently examined model of speech production, that of Kozhevnikov and Chistovich (1965), posits larger units, "articulatory syllables," as the basic units of the speech motor program. The articulatory syllable is described as a CV string, with C being any number of consonants. While this model accounts for some coarticulation data, it completely fails to account for velar function data: the common observation is that the nasality of a consonant is anticipated in a preceding, not following, vowel. Therefore, unless we assume that the organizational units of the motor program are different for different articulators, this model can be eliminated from further consideration.

[4] Binary models are frequently proposed because of their simplicity. However, if a binary model requires a large number of reorganization instructions to account for observational data, it seems that an n-ary model may have equal, or even greater, elegance.

[5] One would expect cumulative velar position as the response of an open-loop system. Such a system would obviate the need for continuous monitoring of velar position, while guaranteeing velopharyngeal port closure adequate for preventing nasal coupling during oral segments.

APPENDIX

A. Preliminaries: Oral Speech

Before considering the acoustic effects of adding the nasal resonator to the pharyngeal and oral resonators, it seems prudent to provide definitions and/or descriptions of concepts that will, of necessity, find their way into the following discussion. This treatment, of course, will not, and could not, be exhaustive.

Traditionally, in evolving an acoustic description of speech, we view the vocal tract as an acoustic tube, one having variable shape and length. For oral speech sounds, the tube is a simple one, having no side branches, with one end at the glottis and the other at the lips.[1] For voiced oral speech sounds, the acoustic properties of such a tube can be described by its _transfer function_, which is the ratio of the volume velocity at the lips to that at the sound source (the glottis). The transfer function can be described by its _poles_, resonances that can be described by their frequencies and bandwidths, or _formants_. The resonance frequencies and their bandwidths are a function of the shape and length of the tube (Fant, 1971; Stevens & House, 1955, 1961). For voiceless speech sounds, the transfer function is the ratio of the volume velocity at the lips to the sound pressure of the source, which, in this condition, is the aperiodic noise or transient excitation generated at the vocal tract constriction (Bell, Fujisaki, Heinz, Stevens, & House, 1961).[2]

B. The Effects of Nasal Coupling

Adding a side branch, or shunt, to the vocal tract tube increases the acoustic complexity of the system in several ways. Among them are the interactions of the poles and _zeroes_ (spectral minima) of the coupled system with those of the "simpler" system. For any given shape of the oral and pharyngeal branches, the transfer function of a system with a coupled side branch (e.g., nasal branch) is determined by the poles and zeroes of the _admittance_ (frequency-dependent susceptibility to flow across a boundary) into the three branches and the pressure gain across each branch (cf. Bell et al., 1961).[3] The pole frequencies of the nasal-branch driving-point admittance (the admittance into the nasal branch, from the velar port) vary with the area of the port, increasing with increasing port size; the zeroes remain fixed. Or, conversely, as the area of the port decreases, the pole frequencies of the nasal-branch driving point admittance decrease, approach the frequencies of their paired zeroes, and are cancelled (Fujimura & Lindqvist, 1971). The poles of the nasal-branch driving-point admittance are the frequencies at which zeroes are observed in the transfer function of the vocal tract; therefore, the closer the pole frequencies of the nasal-branch admittance to the resonances of the rest of the vocal tract, the more extensive will be the effects of adding the nasal branch to the system.

In spite of this complexity, however, it is possible to describe, qualitatively, the results of some of the interactions among the oral, nasal, and pharyngeal resonators. (We will confine ourselves to the effects of nasal coupling on the transfer functions of vowels, and not nasal consonants, because of an interest in understanding observed differences in velar function for different vowels.) First, the lowest formant of the transfer function will fall between the lowest nasal-branch resonance frequency and the lowest formant of the corresponding non-nasalized vowel. More generally, the principal effects of nasal coupling occur in the frequency regions where the admittances into the oral-pharyngeal and nasal branches are most different, particularly in the region of F_1. Nasal coupling also leads to a differential reduction, across vowels, in the amplitude, and an increase in the bandwidth, of F_1, and F_3 is minimized (cf. Fujimura & Lindqvist, 1971; House & Stevens, 1956).

It has not yet been established, however, which one or group of these acoustic effects of nasal coupling has the greatest perceptual salience. Thus, while it is known that close vowels will be perceived as being nasalized at smaller velar port coupling areas than will open vowels (cf. Abramson, Nye, Henderson, & Marshall, 1979; House & Stevens, 1956), we do not know whether the perception of nasality results from amplitude or bandwidth changes, or increasing the center frequency of F_1, or the presence of nasal resonances, or some combination of these or other acoustic results of coupling. Indeed, it may be that the relative positions or intensities of the lowest oral and nasal resonances in the transfer function cue the perception of nasality, especially for small coupling areas that do not have a great overall effect on the vowel spectrum.

FOOTNOTES

[1] This is an overly simplified view, to be sure. It is, however, a useful base for the following discussion. For examples of some of the additional considerations necessary to provide a thorough description or prediction of the acoustic output of the vocal tract, the reader is referred to Fant (1971); Scully and Shirt, (1979); and Stevens (1972).

[2] Complete description of the acoustic output resulting from speech articulation also requires the specification of a radiation function, the ratio of the sound pressure some distance from the lips to the volume velocity at the lips (cf. Fant, 1971; Stevens & House, 1955).

[3] An advance in understanding the interactions between coupled pharyngeal, oral, nasal branches was effected by Mermelstein (Mermelstein, 1971; Rubin, Baer, & Mermelstein, 1979), who established a method for calculating the vocal tract transfer function, based on the independence of the driving point admittances looking into each branch from the velopharyngeal port and of the pressure gain across each branch. This has simplified the techniques necessary for calculating the coupled-system transfer function.

SPEECH PERCEPTION WITHOUT TRADITIONAL SPEECH CUES

Robert E. Remez,+ Philip E. Rubin, David B. Pisoni,++ and Thomas D. Carrell++

Abstract. A three-tone sinusoidal replica of a naturally produced utterance was identified by listeners despite the readily apparent unnatural speech quality of the signal. The time-varying properties of these highly artificial acoustic signals are apparently sufficient to support perception of the linguistic message in the absence of traditional acoustic cues for phonetic segments.

A person listening to a continuously changing natural speech signal perceives a sequence of linguistic elements. Research has attempted to characterize this perceptual process by analyzing the acoustic properties of speech signals that specify the linguistic content (Fant, 1962; Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Mattingly, 1972; Stevens & Blumstein, 1978). In the present study, however, listeners perceived linguistic significance in acoustic patterns with properties differing substantially from those traditionally held to underlie speech perception. And, although listeners accurately reported the linguistic content of these acoustic patterns, the results suggest that the signal was also perceived, simultaneously, to be nonspeech. These novel findings imply that the process of speech perception makes use of time-varying acoustic properties that are more abstract than the characteristic spectra and speech cues typically studied in speech research.

The stimuli used in our study consisted of time-varying sinusoidal patterns that followed the changing formant center-frequencies, the natural resonances of the supralaryngeal vocal tract, of a naturally produced utterance. The sentence, "Where were you a year ago?" was spoken by an adult male talker, digitized at the rate of 10 kHz, and analyzed in sampled data format. Frequency and amplitude values were derived every 15 msec for the center frequencies of the first three formants by the method of linear predictive coding (LPC) (Markel & Gray, 1976). These values were hand-smoothed in some portions to ensure continuity, and were used as synthesis parameters for a digital sinewave synthesizer. Three time-varying sinusoids were then generated to match the LPC-derived center frequencies and amplitudes of the first three formants, respectively, of the natural speech utterance. Figure 1 shows

---

+Barnard College, Columbia University.
++Indiana University.

Acknowledgment. We gratefully acknowledge the assistance of Charles Marshall, Joe Montelongo, and Sue Gans. We also thank Dick Aslin, Peter Bailey, Alvin Liberman, and Frank Restle, among many others, for very helpful advice and comments on an earlier version of the paper. This research was supported by NIMH Grant MH 32848 (to Robert E. Remez), NIMH Grant MH 24027 (to David B. Pisoni) and NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

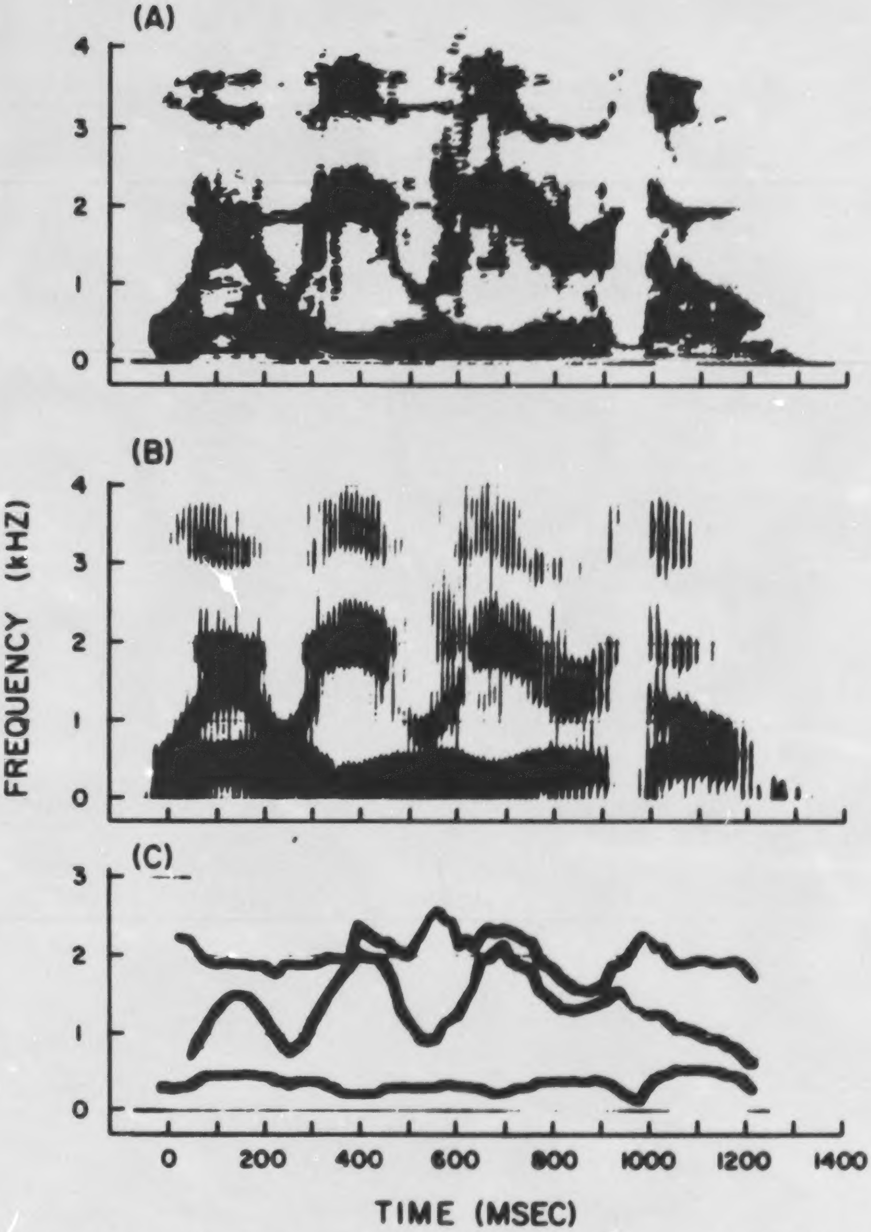


Figure 1. (a) Narrowband spectrogram of the natural utterance, "Where were you a year ago?" showing harmonic structure as narrow horizontal lines along the frequency scale. (b) Wideband spectrogram of the same utterance, showing formant pattern as dark bands along the time axis. Note that the vertical striations correspond to individual laryngeal pulses. (c) Narrowband spectrogram of the three-tone sinusoidal replica. The energy concentrations follow the time-varying pattern of the formants above, but there is no energy present except at the format center frequencies. The figure does not accurately reproduce the amplitude variation in the sinusoidal pattern.

narrowband and wideband spectrograms of the original spoken utterance and a narrowband spectrogram of its replica formed by the three time-varying sinusoids.

Although our synthetic stimuli were designed to preserve the frequency and amplitude variation of natural speech formants, the three-tone patterns differ from natural speech in several prominent ways. First, the energy spectra of the tones differ greatly from those of natural and synthetic speech. Voiced speech sounds, produced by pulsed laryngeal excitation of the supralaryngeal cavities, exhibit a characteristic spectrum of harmonically related values (Chiba & Kajiyama, 1941; Fant, 1960) [1]. Because the frequencies of the individual tones in our stimuli follow the formant center frequencies, the components of the spectrum at any moment are not necessarily related as harmonics of a common fundamental. In essence, the three-tone pattern does not consist of harmonic spectra, although natural voiced speech does.

Second, the short-time spectra of the tone stimuli lack the broadband formant structure that is also characteristic of speech (including whispered speech). Because the resonant properties of the supralaryngeal vocal tract introduce short-time amplitude maxima and minima across the harmonic spectrum of energy generated at the larynx, some frequency regions contain harmonics with more energy than neighboring regions [2]. Our tone stimuli consist of no more than three sinusoids, and therefore no energy is present in the spectrum except at the particular frequencies of each tone. Thus, the short-time spectra of the tone stimuli are also distinct in this way from the energy spectra of natural speech. There is literally no formant structure to the three-tone complexes, though the tones do exhibit acoustic energy at frequencies identical to the center frequencies of the formants of the original, natural utterance.

Third, the dynamic spectral properties of speech and tone stimuli are quite different. Across phonetic segments the relative energy of each of the harmonics of the speech spectrum changes. Formant center-frequencies may be computed by following the changes in amplitude maxima of the harmonic spectrum. However, natural speech signals do not exhibit continuous formant frequency variation. Rather, laryngeal activity in voiced speech creates distinct pulses characterized by a formant structure. Thus, changes in formant structure, particularly when observed in wideband spectrograms, may erroneously appear to contain continuous formant variation over time. Figure 1b displays a wideband spectrogram, in which the finegrained amplitude differences are averaged over frequency to derive the formant pattern. In contrast to the case in speech, each tone in our stimuli continuously follows the computed peak of a changing resonance of the natural utterance. Overall, our three-tone pattern is a deliberately abstract representation of the time-varying spectral changes of the naturally produced utterance, though in local detail it is unlike natural speech signals.

The complex tone signal, having neither fundamental period nor formant structure, consists of none of those distinctive acoustic attributes that are assumed traditionally to underlie speech perception. None of the appropriate acoustic cues based on the acoustic events within speech signals is present in our stimuli, for example, neither formant frequency transitions, which cue manner and place of articulation; nor steady state formants, which cue vowel color and consonant voicing; nor fundamental frequency changes, which cue voicing and stress (Liberman & Studdert-Kennedy, 1978). Similarly, the short-time spectral cues, which depend on precise amplitude and frequency characteristics across the harmonic spectrum, are absent from these tonal stimuli, for example, the onset spectra that are often claimed to underlie perception of place features (Stevens & Blumstein, in press). The perceptual importance of these attributes of speech signals has been rationalized by theoretical models of sound production in the vocal tract. These models describe the speech signal as the product of a source and a filter (Chiba & Kajiyama, 1941; Stevens, 1964). Briefly, glottal pulsing provides a source in which energy is present at integral multiples of the fundamental frequency. The complex resonances of the pharyngeal, oral and nasal cavities of the vocal tract are treated as a time-varying filter; the peaks in the vocal-tract transfer function represent the formants. Perceptual tests of potentially distinctive attributes, however, have typically employed electronic or digital analogs of the source-filter theory of speech acoustics to create stimuli. In doing so, these tests have not questioned the necessity of harmonic spectra or broadband formant structure in speech perception; nor have they empirically raised the possibility that listeners attend to higher-order relational properties of time-varying speech signals.

The present study is a test of these assumptions. The absence of traditional acoustic cues to phonetic identity suggests that our sinusoidal replica of the sentence should be perceived to be three independently changing tones. However, if listeners are able to perceive the tones as speech, then we may conclude that traditional speech cues are themselves approximations of second-order signal properties to which listeners attend when they perceive speech.

Our perceptul test consisted of three conditions in which independent groups of listeners were informed to different degrees about the tonal stimuli that they would hear [3]. Within each instructional condition, different groups of eighteen listeners each were assigned to seven stimulus conditions: the three tones presented together (S1:T1+T2+T3); three pairwise tone combinations (S2:T1+T2; S3:T2+T3; S4:T1+T3); and each tone played separately (S5:T1; S6:T2; S7:T3). The three instructional conditions crossed with the seven stimulus conditions made twenty-one experimental conditions in all. In each condition a given sinusoidal pattern was presented four times in succession, at approximately 85 dB SPL, by audiotape playback over matched and calibrated headphones.

In Instructional Condition A, listeners were asked simply to report their spontaneous impressions of the stimuli, having been told nothing in advance of the nature of the sounds. Multiple responses were permitted. The accumulated responses, organized by stimulus condition, are displayed in Table 1. Apparently, the presentation of tones following the formant center-frequencies is insufficient to elicit phonetic perception; modal responses in each stimulus condition indicate that the majority of listeners did not hear the sinusoids as speech. A small number of responses in several conditions favored human- or artificial-speech interpretations, though, and two listeners in the three-tone condition responded that they heard the sentence, "Where were you a year ago?" This outcome might be anticipated only if there were

Table 1

Response Categories and Frequencies by Stimulus Condition in Instructional Condition A

STIMULUS CONDITION	RESPONSE CATEGORIES
S1 (T1+T2+T3)	Science fiction sounds (8), Computer bleeps (5), Music (4), Several simultaneous sounds (3), Human speech (3), Where were you a year ago (2), Radio interference (2), Human vocalizations (1), Artificial speech (1), Bird sounds (1), Reversed speech (1)
S2 (T1+T2)	Science fiction sounds (7), Computer bleeps (3), Sirens (2), Music (2), Radio interference (2), Tape recorder problems (1), Reversed speech (1), Whistles (1), Artificial speech (1), Human speech (1)
S3 (T2+T3)	Science fiction sounds (14), Radio interference (3), Music (2), Computer bleeps (2), Whistles (1), Several simultaneous sounds (1)
S4 (T1+T3)	Science fiction sounds (9), Artificial speech (5), Computer bleeps (4), Several simultaneous sounds (4), Whistles (3), Radio interference (2), Tape recorder problems (2), Human speech (1), Human vocalizations (1), Reversed speech (1), Music (1)
S5 (T1)	Science fiction sounds (5), Music (4), Reversed speech (4), Tape recorder problems (3), Human speech (2), Artificial speech (2), Animal cries (2), Bird sounds (2), Radio interference (2), Several simultaneous sounds (2), Human vocalizations (1)
S6 (T2)	Sirens (7), Bird sounds (6), Mechanical sound effects (4), Radio interference (4), Animal cries (3), Whistles (2), Computer bleeps (1)
S7 (T3)	Bird sounds (17), Whistles (6), Mechanical sound effects (5), Human vocalizations (3), Human speech (1), Artificial speech (1), Computer bleeps (1), Animal cries (1), Music (1), Radio interference (1), Tape recorder problems (1)

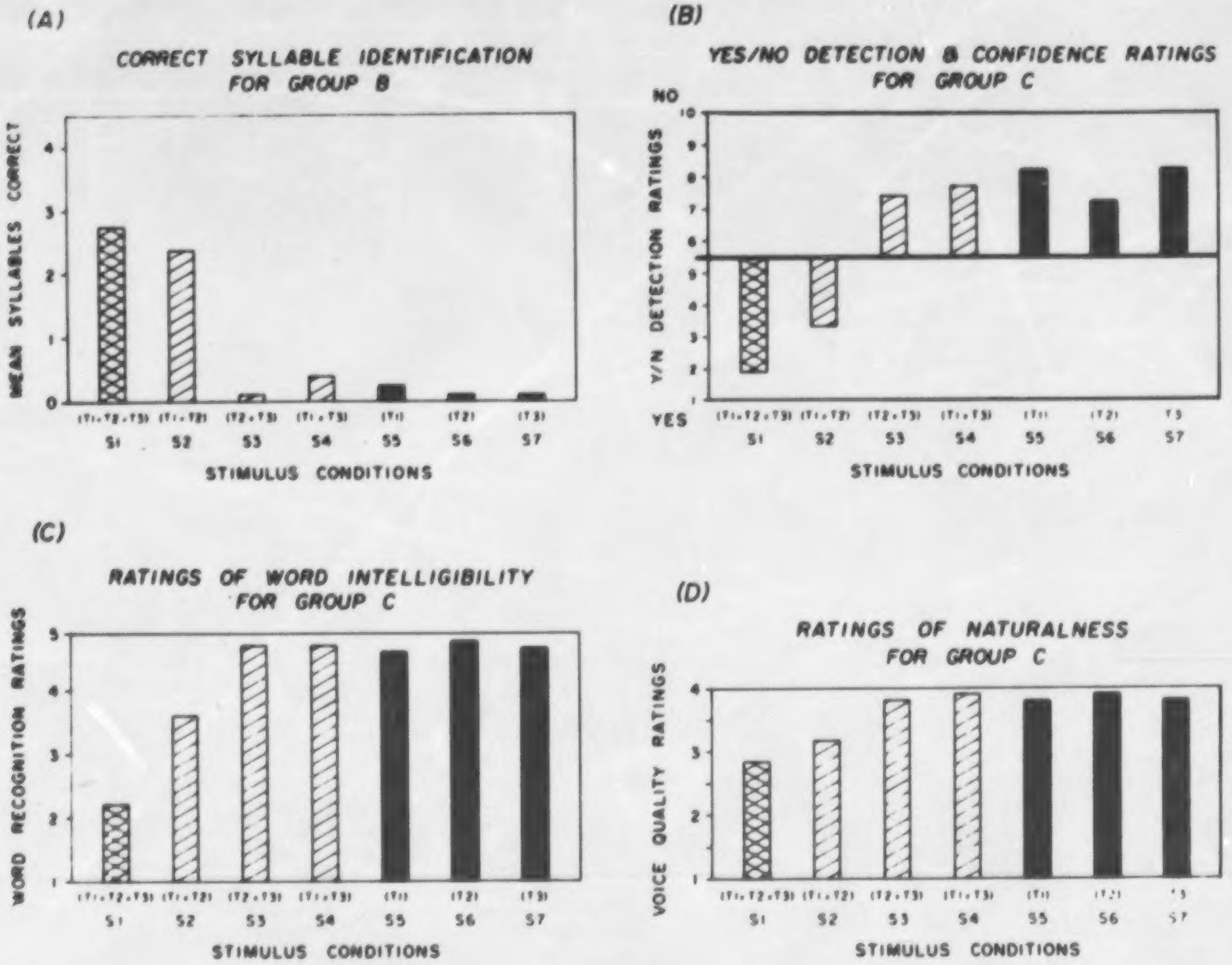
stimulus support of some kind for perceiving the linguistic content of these patterns. Even as a response to a direct request to generate a sentence in English, the probability of producing this exact sentence is exceedingly small (Miller & Chomsky, 1960).

In Instructional Condition B, listeners were informed that they would hear a sentence produced by a computer, and were asked to transcribe the synthetic utterance as faithfully as possible. We scored the responses in each condition for correct number of syllables transcribed relative to the original utterance, "Where were you a year ago?" Average transcription performance in each stimulus condition is presented in Figure 2a. It is clear that a large number of subjects can identify the sentence in Conditions S1 and S2. Nine of the listeners across these two conditions transcribed the entire sentence correctly, though ten others reported that they could hear no sentence at all in the tones. The remaining listeners transcribed various syllables correctly. We conclude from these first two instructional conditions that naive listeners may not automatically perceive sinusoidal replicas of natural speech as linguistic entities. When instructed to do so, however, they perform well presumably because the linguistic information, though not carried by acoustic elements producible by a vocal tract, is preserved in the time-varying relational structure of the stimulus pattern [4].

In Instructional Condition C, listeners were asked directly to evaluate the speech quality of the tone stimuli. They were told that they would be presented with the sentence, "Where were you a year ago?" and they were asked to make three judgments. First, they reported whether the sentence was discernible in the tonal pattern by responding Yes or No; they also provided a confidence rating for their judgments using a dual five-point scale. These responses were converted to a ten-point scale (1=confident Yes; 10=confident No). The scores are presented in Figure 2b grouped by stimulus condition. In five of the stimulus conditions, listeners were very confident that they did not hear the sentence in the tones. However, in Conditions S1 and S2, listeners were very confident that they recognized the intended sentence; the average confidence ratings in these two conditions did not differ significantly despite the absence of Tone 3 in Condition S2 (Scheffe post hoc means test, $p>.1$).

In the second task, listeners rated the number of words that could be identified in the particular pattern presented (1=all, 2=most, 3=a few, 4=almost none, 5=none). As shown in Figure 2c, for five of the stimulus conditions subjects indicated that they could not identify any of the words in the sentence. But, in the three-tone condition (S1), listeners reported that almost every word was clear. The omission of Tone 3 from the pattern in Condition S2 led subjects to report that significantly fewer words were intelligible (Scheffe test, $p<.025$), yet this condition remains significantly different from Conditions S3 through S7 (Scheffe test, $p<.001$).

In the third task, listeners rated the voice quality of the tone stimuli [1=natural, 2=funny (peculiar), 3=unnatural, 4=nonspeech]. The average ratings appear in Figure 2d. The split between S1 and S2 and the other conditions is still quite evident, as it was in Condition B above; however, we see here that these two stimulus patterns were judged to have unnatural voice quality despite their clear intelligibility. In essence, listeners apprehend


(A) CORRECT SYLLABLE IDENTIFICATION FOR GROUP B

(B) YES/NO DETECTION & CONFIDENCE RATINGS FOR GROUP C

(C) RATINGS OF WORD INTELLIGIBILITY FOR GROUP C

(D) RATINGS OF NATURALNESS FOR GROUP C

gure 2. (a) Transcription performance for Instructional Condition B. (b) Detection ratings for Instructional Condition C (1=Confident Yes, 10=Confident No); (c) Ratings of number of intelligible words in the tones (1=every, 2=most, 3=a few, 4=almost none, 5=none); (d) Naturalness ratings (1=natural, 2=peculiar, 3=unnatural, 4=nonspeech). Cross hatched=three-tone stimulus; hatched=two-tone stimulus; filled= single-tone stimulus.

the linguistic significance of the tonal patterns despite the radically unnatural, nonspeech quality [5,6]. That is, they were able to perceive the linguistic content of the utterance in the absence of acoustic patterns of the kind generated by the human vocal tract

The results of the present study cannot be explained within the framework of existing theories of speech perception [7], for the tones contained none of the elemental acoustic cues typically held to underlie speech perception (i.e., formant structure, fundamental period, or distinctive short-time spectra). Though the tones present information about formant center-frequency, this minimal structure is evidently not sufficient to elicit phonetic perception spontaneously, as we saw in the performance of the naive listeners in Condition A. In fact, no property of the three-tone stimulus obliges the listener to hear it phonetically--except that its time-varying pattern of frequency change corresponds abstractly to the potential acoustic products of vocalization [8]. The linguistically primed listeners in Conditions B and C are capable, for the most part, of directing their attention to the phonetic properties of the sinusoidal signal, merely by virtue of the instruction to listen in the "speech mode" of perception. For these subjects, the tones provide sufficient stimulation to evoke phonetic perception, albeit a kind that also identifies the "vocal" source as unnatural. We conclude, then, that speech perception can endure the absence of particular short-time acoustic spectra and traditional formant-based acoustic cues only insofar as the pattern of change in the natural signal is preserved over transposition from harmonic to sinewave spectra [9]. Further examples of nonspeech tonal analogues of natural speech utterances are needed to characterize more precisely the time-varying relations within the acoustic patterns that support phonetic perception.

REFERENCES

Allen, J. Synthesis of speech from unrestricted text. *Proceedings of the IEEE*, 1976, *64*, 433-442.

Bailey, P. J., Summerfield, A. Q., & Dorman, M. On the identification of sine-wave analogues of certain speech sounds. *Haskins Laboratories Status Report on Speech Research*, 1977, *SR-51/52*, 1-25.

Best, C. T., Morrongiello, B., & Robson, R. The perceptual equivalence of two acoustic cues for a speech contrast is specific to phonetic perception. *Perception & Psychophysics*, in press.

Boring, E. G. *Sensation and perception in the history of experimental psychology*. New York: Holt, 1942.

Chiba, T., & Kajiyama, M. *The vowel: Its nature and structure*. Tokyo: Tokyo-Kaiseikan Publishing Company, 1941.

Cutting, J. E. Two left-hemisphere mechanisms in speech perception. *Perception & Psychophysics*, 1974, *16*, 601-612.

Fant, C. G. M. *The acoustic theory of speech production*. The Hague: Mouton, 1960.

Fant, C. G. M. Descriptive analysis of the acoustic aspects of speech. *Logos*, 1962, *5*, 3-17.

Grunke, M. E., & Pisoni, D. B. Perceptual learning of mirror-image acoustic patterns. In E. Fischer-Jorgenson, J. Richel, & N. Thorsen (Eds.), Proceedings of the Ninth International Congress of Phonetic Sciences, Vol. II. Copenhagen: Institute of Phonetics, 1979, 461-467.

Hermann, L. Nachtrag zur Untersuchung der Vocalcurven. Archiv für die Gesamte Physiologie des Menschen und der Tiere, 1894, 58, 264-269.

Köhler, W. Akustiche Untersuchungen, II. Zeitschrift für Psychologie mit Zeitschrift für Angewandte Psychologie und Charakterkunde, 1910, 58, 59-140.

Kuhn, G. M. Stop consonant place perception with single-formant stimuli: Evidence for the role of the front-cavity resonance. Journal of the Acoustical Society of America, 1979, 65, 774-788.

Liberman, A. M., & Cooper, F. S. In search of the acoustic cues. In A. Valdman (Ed.), Papers in linguistics and phonetics to the memory of Pierre Delattre. The Hague: Mouton, 1972, 329-338.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 421-461.

Liberman, A. M., Mattingly, I. G., & Turvey, M. T. Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), Coding processes in human memory. Washington, D.C.: V. H. Winston, 1972, 307-334.

Liberman, A. M., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. Leibowitz, & H.-L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII, Perception. New York: Springer, 1978, 143-178.

Markel, J. D., & Gray, A. H., Jr. Linear prediction of speech. New York: Springer Verlag, 1976.

Mattingly, I. G. Speech cues and sign stimuli. American Scientist, 1972, 60, 327-337.

Miller, G. A., & Chomsky, N. Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology (Vol. II). New York: Holt, 1960, 419-491.

Modell, J. D., & Rich, G. J. A preliminary study of vowel qualities. American Journal of Psychology, 1915, 26, 453-456.

Stevens, K. N. Acoustical aspects of speech production. In W. O. Fenn & H. Rahn (Eds.), Handbook of physiology-respiration I. Washington, D.C.: American Physiological Society, 1964, 347-355.

Stevens, K. N. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. David & P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw-Hill, 1972, 51-66.

Stevens, K. N., & Blumstein, S. E. Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America, 1978, 64, 1358-1368.

Stevens, K. N., & Blumstein, S. E. The search for invariant acoustic correlates of phonetic features. In P. D. Eimas & J. L. Miller (Eds.), Perspectives in the study of speech. Hillsdale, N.J.: Lawrence Erlbaum, in press.

FOOTNOTES

[1] The closely spaced horizontal lines shown in Figure 1a are the harmonics of the fundamental frequency of phonation, and are typically revealed in narrowband spectrograms.

[2] Typically, the amplitude of the valleys in the spectrum of natural speech ranges from 10-30 dB below the amplitude of the peaks (Stevens & Blumstein, in press).

[3] Our listeners were students of introductory psychology at Indiana University in Bloomington. They were naive with respect to synthetic speech.

[4] It has often been emphasized that a variety of acoustic events may cue a single phonetic feature in the absence of other, redundant cues; experiments with synthetic speech in which phonetic distinctions were minimally cued indicate that listeners tolerate schematized speech signals with little loss of intelligibility (Liberman & Cooper, 1972). For this reason, listeners probably do not require stimuli to display the acoustic "stigmata" of speech to be candidates for phonetic interpretation (Liberman, Mattingly, & Turvey, 1972). However, even schematized synthetic speech has consisted of acoustic cues that are utterable in principle as components of a speech signal; these cues enjoy specific articulatory rationales. This resemblance of schematized synthetic speech to natural speech may have led theorists to underestimate the abstractness of the stimulus properties relevant to perception. Signals consisting of sinusoids may be used to study these more abstract, time-varying acoustic properties underlying phonetic perception, for their phonetic effects can neither be explained by arguing that they are components of natural signals; nor by arguing that they are acoustic products of vocal articulation.

[5] Although much intelligible synthetic speech would also be judged unnatural, this may be ascribed to the practice of presenting the speech cues in contexts of minimal variation in the acoustic parameters that are irrelevant to intelligibility—which affect speech quality nonetheless (Liberman & Cooper, 1972). A synthesizer that produces a harmonic spectrum, broadband formants and a fundamental period within the normal range will sound unnatural, and perhaps be unintelligible, despite the acoustic resemblance to natural speech if the synthesis of prosodic variation--of speech rhythm, meter, and melody—is inappropriate (Allen, 1976). The judgment that this kind of synthetic imitation of speech signals is unnatural is, therefore, quite different from the judgment of unnaturalness in the present case.

[6] Although the intelligibility of our sinusoidal sentence is predicted by the co-occurrence of T1 and T2, but not of T1 and T3, the effectiveness of each tone pair will vary as a function of the phonetic composition of the utterance. While the resonance associated with the oral cavity is primary in its importance for phonetic perception (Kuhn, 1979), either F2 or F3 may be affiliated with the oral cavity, depending on the phone in question (Stevens, 1972). Therefore, the critical tone pair will sometimes include T2, sometimes T3, depending on the phonetic composition of the utterance.

[7] The proposal that listeners "track" formant frequency variations must be entertained as an explanation of our findings only if the meaning of the term "formant" is extended to mean "any peak in the spectrum." In its present sense the concept of the formant refers to a natural resonance of the vocal cavities (Hermann, 1894). Quite literally, then, there are no vocal resonances in our tone complexes (though listeners who succeed in extracting the meaning of the "utterance" probably do so because the tones preserve time-varying properties of vocally produced signals). Our preference is to retain the literal meaning of "formant," and to conclude, therefore, that the difference between voiced speech signals and the tone signals is that the former contain broadband formant structure and harmonic spectra, and the latter merely inharmonic peaks with infinitely narrow bandwidths.

[8] Our finding is related, in some sense, to early studies of "vowel pitch" in which simple steady state tones were judged to possess "vocality," or speechlike qualities (Köhler, 1910; Modell & Rich, 1915; Titchener [described in Boring, p. 374, 1942]). More recent studies have shown that listeners may identify brief complex sinusoidal patterns as isolated syllables, and therefore as speech sounds, when they are supplied with restricted response alternatives in low uncertainty judgment tasks (Cutting, 1974; Bailey, Summerfield & Dorman, 1977; Best, Morrongiello & Robson, in press; Grunke & Pisoni, 1979). The present study, however, makes use of neither a closed response set nor a low uncertainty task to obtain the effect of intelligibility.

[9] We have recently synthesized the sentence, "A yellow lion roared," thereby extending the range of tone synthesis to nasal manner as well as the stop consonant, liquid consonant, and vowel phone classes represented here. Similar findings have been obtained with this sentence, indicating that the present results are not due to peculiarities of the sentence used in these tests.

INFLUENCE OF PRECEDING LIQUID ON STOP CONSONANT PERCEPTION*

Virginia A. Mann+

Abstract. Certain attributes of a syllable-final liquid can influence the perceived place of articulation of a following stop consonant. To demonstrate this perceptual context effect, the CV portions of natural tokens of [al-da], [al-ga], [ar-da] and [ar-ga] were excised and replaced with closely matched synthetic stimuli drawn from a [da]-[ga] continuum. The resulting hybrid disyllables were then presented to listeners who labeled both liquids and stops. The natural VC portions had two different effects on perception of the synthetic CVs. First, there was an effect of liquid category: Listeners perceived "g" more often in the context of [al] than in that of [ar]. Second, there was an effect due to tokens of [al] and [ar] having been produced before [da] or [ga]: More "g" percepts occurred when stops followed liquids that had been produced before [g]. Spectrograms of the original utterances indicate that each of these perceptual effects finds a parallel in speech production. Here, it seems, is another instance where speech perception compensates for coarticulation during speech production.

When an utterance is articulated, the gestures for adjacent phones overlap and become interwoven. One consequence of this coarticulation is that stop consonants may have slightly different places of occlusion when they occur in different phonetic sequences. To date, the best-known illustration of this point concerns the shift in place of occlusion that is consequent upon a change in the preceding or following vowel. Velar stops receive a more forward place of occlusion when they are adjacent to a front vowel such as [i] than when they are adjacent to a back vowel such as [a] (Gay, 1977; Öhman, 1966). Another example, which has recently emerged from Repp and Mann's (in press) perceptual and acoustic observations of stops in fricative-stop clusters, is that when [t] or [k] follow [s], these stops can receive a relatively more forward place of articulation than when they follow [ʃ].

Insofar as coarticulation with adjacent phones causes shifts in the place of stop occlusion and, correspondingly, changes in the acoustic signal that

---

*To be published in Perception & Psychophysics.

+Also at Bryn Mawr College.

Acknowledgment. This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories, and by NICHD Postdoctoral Fellowship HD05677 to the author. Some of the results were reported at the 99th meeting of the Acoustical Society of America in Atlanta, April 1980. I would like to thank Bruno Repp and Alvin Liberman for their advice at all stages of this project.

reflect stop production, we should suppose that perception of a stop consonant must often require the integration of acoustic cues that are numerous, diverse and context-sensitive. That listeners do, in fact, integrate such cues in the process of stop perception can be seen in the existence of two perceptual "context effects" that reflect perceptual compensation for the particular coarticulatory effects cited above. With regard to the relative fronting of velar stops before vowels such as [i]--which causes release bursts to be relatively higher in frequency--Liberman, Delattre, and Cooper (1952) have shown that when steady-state synthetic vowels are preceded by bursts of various frequencies, listeners require a higher-frequency burst to hear [k] before [i] than before [a] (Liberman et al., 1952). With regard to the fronting of stops following [s], Mann and Repp (in press-a) report that when stimuli from a [ta]-[ka] continuum are preceded by a fricative noise appropriate to [s], listeners give more "k" responses than when the preceding noise is appropriate to [ʃ].

These and other instances where perceptual findings parallel the dynamics of speech production have led some investigators (e.g., Liberman et al., 1952; Mann & Repp, in press-a, in press-b; Repp, Liberman, Eccardt, & Pesetsky, 1978; Repp & Mann, in press) to the view that speech perception operates with reference to the dynamics of speech production. According to this view, perceptual context effects in stop perception should be found wherever stop production is influenced by production of an adjacent phonetic segment. This prediction is clearly upheld by the above-mentioned findings that stop perception is influenced by an adjacent vowel (Liberman et al., 1952) or fricative (Mann & Repp, in press-a). The purpose of the present experiment was to determine whether perceived place of stop occlusion could be influenced by a preceding liquid, since it seemed possible that a preceding liquid can influence the production of a following stop.

There are two circumstances under which a liquid may precede a stop: The liquid and stop may either occur as a syllable-final cluster, or be separated by a syllable boundary. Here I have focused on liquid-stop sequences of the latter type, since in that case a finding that liquids influence stop perception would have the additional implication that listeners are able to integrate perceptual information across a syllable boundary. One might expect the preceding liquid to influence perception of the following stop in a disyllable such as [al-da], since articulation of the liquid most probably overlaps that of the stop. Although the literature does not provide any systematic observations on liquid-stop clusters, it seems at least possible that stops that follow [l] may receive a more forward place of articulation than those that follow [r], considering the fact that coarticulatory effects tend to be assimilatory in nature. It further seems highly likely that the place of stop occlusion is reflected in the portion of the utterance immediately preceding the closure (i.e., in the portion commonly associated with the liquid). Thus, there might be coarticulatory effects in both directions, with appropriate acoustic and perceptual consequences.

The present experiment addressed these possibilities by excising naturally-produced VC syllables from utterances of [al-da], [al-ga], [ar-da] and [ar-ga] and following them with stimuli from a synthetic [da]-[ga] continuum. Two questions were of interest: First, would a preceding [l] lead to more "g" responses than a preceding [r]? If so, it would suggest that listeners compensate in perception for a "left-to-right" coarticulatory influence of the liquid on the stop. Second, would liquids that had been coarticulated with [ga] lead to more "g" percepts than those coarticulated with [da]? If so, it would suggest that listeners are sensitive to a "right-to-left" coarticulatory influence of the stop on the liquid. In addition, as a means of obtaining more direct evidence for the coarticulatory phenomena underlying the two proposed perceptual effects, acoustic measurements were made of the utterances from which the stimuli were constructed.

EXPERIMENT

Method

Subjects. The subjects included the author, a research assistant, and eight paid volunteers. As experience with listening to synthetic speech did not seem to influence the pattern of results, all data were pooled.

Materials. A male, phonetically-trained native speaker of English (LJR) produced six repetitions each of [al-da], [al-ga], [ar-da], and [ar-ga]. These disyllables were produced according to a random sequence in which, as a control for any effects of stress pattern, half received syllable-initial stress and half received syllable-final stress. All utterances were recorded onto magnetic tape, using a Shure dynamic microphone in a soundproof room, before being digitized at 10,000 Hz using the Haskins Laboratories Pulse Code Modulation (PCM) System. Subsequently, separate files were created for the VC and CV portions of each disyllable, i.e., the signal portions preceding and following the stop closure interval. The VC syllables were stored for later use in constructing "hybrid" disyllables. Their durations and relative peak amplitudes are listed in Table 1. The natural CV syllables were analyzed, using the CONVERT program in conjunction with the Haskins Laboratories OVE IIIc synthesizer. (See Kuhn, 1977, for details of the CONVERT procedure.) Their duration, pitch contour, amplitude contour, and average formant frequencies were taken as guidelines for constructing two seven-member [da]-[ga] continua. The stimuli along each continuum differed only in the onset of F_3, which ranged from 2690 to 2104 Hz in approximately equal steps. Onset values for F_1 and F_2 transitions were fixed at 310 and 1588 Hz, respectively. Steady-state values for the first three formants were 649, 1131, and 2448 Hz, respectively and all formant transitions were stepwise linear and 100 msec in duration. For stimuli along the "stressed" continuum, stimulus duration (240 msec), amplitude contour, and pitch contour were those of a syllable (chosen at random from the several tokens) that had received primary stress. For those along the "unstressed" continuum, duration (180 msec), amplitude contour and pitch contour were those of a syllable (also chosen at random) that had not been stressed. The relative peak amplitude of the "unstressed" syllables was 3 dB below that of the "stressed" syllables. The two continua were otherwise identical, with each stimulus from the stressed continuum having the same formant structure as the corresponding stimulus from the unstressed continuum.

Table 1

Mean Duration and Intensity for Naturally-Produced VC Syllables
(Standard deviations in parentheses)

	[al-(da)]	[al-(ga)]	[ar-(da)]	[ar-(ga)]
Duration (msec)				
V́C-CV	278(24)	252(29)	287(14)	248(22)
VC-CV́	240(3)	245(9)	239(11)	243(13)
Relative peak amplitude (dB, arbitrary reference)				
V́C-CV	9.1(0.4)	9.4(1.0)	5.1(2.8)	6.4(0.8)
VC-CV́	-6.0(1.3)	-9.0(1.3)	-3.6(0.1)	-5.3(1.6)

The actual test materials were constructed by combining the previously stored natural VC syllables with the stimuli along the two synthetic continua. All synthetic stimuli were first digitized at 10,000 Hz; stimuli along the stressed continuum were then preceded by tokens of [al] and [ar] that had not received primary stress, whereas stimuli along the unstressed continuum were preceded by VC tokens that had received primary stress. In all cases, a 50-msec silent gap separated VC offset from the onset of the synthetic CV syllable. This value, although slightly shorter than the mean closure duration of the original natural utterances (80 msec), was still within the range of closure durations found in those utterances. As there were 12 tokens of [al] and 12 of [ar] (3 tokens, 2 contexts, and 2 stress conditions), combination of each token with the seven stimuli from along the appropriate synthetic continuum resulted in a total of 168 hybrid disyllables. These disyllables were recorded onto a test tape (the VC-CV tape) in two randomized sequences, with interstimulus intervals of 3 sec and longer pauses between sets of 56 stimuli. A second test tape (the CV tape) contained a randomized sequence of the stimuli along the two [da]-[ga] continua, repeated twelve times.

Procedure. Each subject participated in a single eighty-minute session during which he or she was seated in a soundproof room listening to stimuli over TDH-39 earphones. The CV tape was presented first, followed by a short break. There was next a short practice sequence of hybrid disyllables that contained only the endpoint stimuli from the two CV continua; it was followed by two presentations of the VC-CV test tape. Thus, each stimulus was presented 12 times (ignoring token differences in the natural-speech portions).

In responding to the CV tape, the subjects were asked to identify each stop as "d" or "g". For the hybrid disyllables, they were asked both to identify the liquid as "l" or "r" and the following stop as "d" or "g".

Results

The procedure of combining natural and synthetic syllables into single test utterances was highly successful. In fact, several listeners spontaneously praised the disyllables' resemblance to natural speech. None of the subjects had any difficulty hearing both liquids and stops; moreover, all of them were completely accurate in labeling the liquid consonants.

Consider first the pattern of responses to the isolated CV stimuli. Figure 1 plots the percentage of "g" responses given to each stimulus as a function of F3 onset frequency. It can be seen that stimulus 1, which contained a third-formant onset frequency appropriate for [da], received no "g" responses, while stimulus 7, which contained a third-formant onset frequency appropriate for [ga], received 100 percent "g" responses. Between these two endpoints, the function follows the ogive pattern characteristic of identification functions obtained with stop consonant continua. Note that the function obtained with stimuli whose duration, pitch contour, and amplitude contour were appropriate for a CV in stressed position (dashed line) is no different from that obtained with stimuli whose structure was appropriate for a CV in unstressed position (solid line).

Let us now turn to the main concern of this study, which was the question of whether labeling of stimuli along the [da]-[ga] continua would be altered by the presence of a preceding liquid. In the introduction, two possible effects were outlined, one concerning an effect of liquid category, the other concerning an effect due to the liquids having been produced before [d] or [g]. The effect of liquid category membership was hypothesized to be that a preceding [l] would, in general, lead to more "g" responses than a preceding [r]. The relevant results are graphed in Figure 2, where it can be seen that the hypothesis was confirmed. There is a clear difference between the effects of preceding [l] (solid line) and preceding [r] (dashed line): Stops preceded by [l] were much more likely to be assigned a velar place of articulation. This effect was highly significant, $F(1,9) = 52.16$, $p < .0005$, and primarily due to [l]: There was no significant difference between the percentage of "g" responses given to CV stimuli preceded by [r] and that for CV stimuli presented in isolation, but labeling of stimuli preceded by [l] significantly differed from the baseline, $F(1,9) = 50.1$, $p < .0005$. A comparison of the left and right panels of Figure 2 further reveals that the difference between the effects of [l] and [r] on stop perception was somewhat greater when the syllable containing the liquid did not receive primary stress, $F(1,9) = 8.13$, $p < .025$. However, this paradoxical effect of stress did not appear to hold for all individual tokens of [al] and [ar], since it fell short of significance in a minF' analysis (Clark, 1973). The effect of liquid context, on the other hand, remained significant, $\mathrm{min}F'(1,5) = 18.4$, $p < .01$.

The second question asked in the introduction was whether tokens of [al] and [ar] that had been produced before [ga] would lead to more "g" responses than those produced before [da], all other things being equal. In that case,

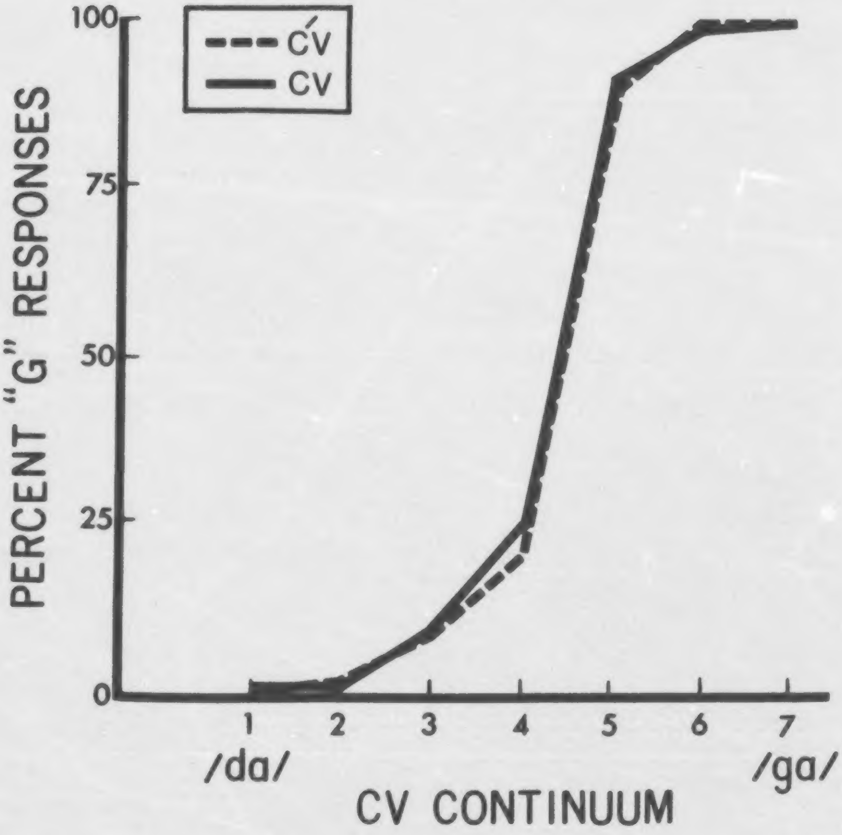


Figure 1. Percentage of "g" responses given to isolated CV stimuli from the synthetic [da]-[ga] continua.

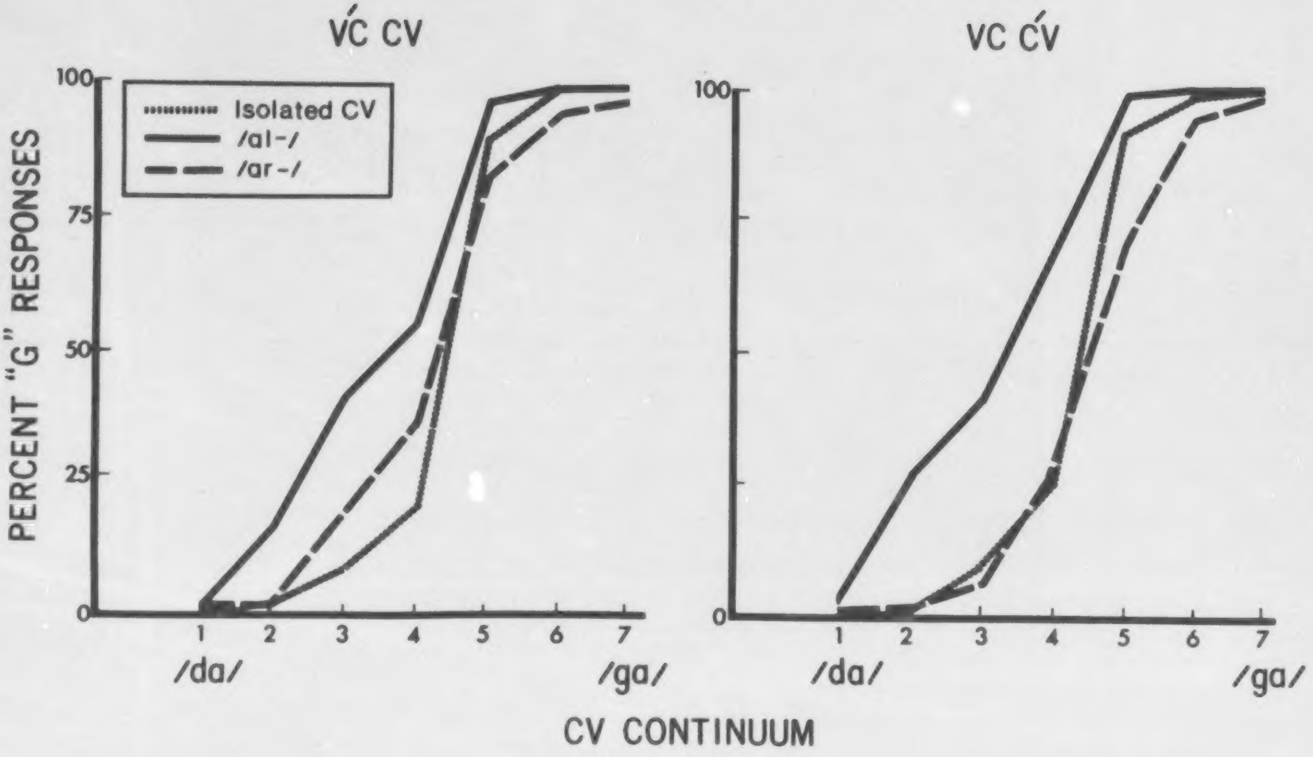


Figure 2. Percentage of "g" responses given to CV stimuli as a function of the category of the preceding liquid.

the relevant results are graphed in Figure 3, where the left panel shows the percentage of "g" responses to synthetic CV stimuli preceded by [al] and the right panel shows the corresponding percentages for stimuli preceded by [ar]. In each panel it can be seen that liquids that had been produced before [ga] (dashed line) led to more "g" responses than those produced before [da] (solid line). It is further evident that the effect is considerably stronger for [ar] than for [al]. An analysis of variance computed on the percentage of "g" responses reveals a significant effect of original stop ([g] vs. [d]), $F(1,9) = 35.63$, $p < .0005$, and an interaction between this effect and liquid category, $F(1,9) = 13.32$, $p < .005$. Neither of these effects was influenced by the stress pattern of the disyllables, and both are upheld by the results of a minF' analysis with tokens treated as a random variable. For the effect of original stop, $\mathrm{min}F'(1,11) = 28.0$, $p < .0005$; for the interaction between this effect and liquid category, $\mathrm{min}F'(1,7) = 6.74$, $p < .05$.

Discussion

Through a technique of combining natural and synthetic syllables into hybrid disyllables, the present experiment revealed that certain attributes of a preceding liquid can influence the perceived place of stop occlusion. Two influences are evident in the pattern of stop labeling functions obtained when naturally-produced tokens of [al] and [ar] preceded stimuli along a [da]-[ga] continuum. First, there was an influence of liquid category: Many more "g" percepts occurred when synthetic CV stimuli were preceded by [l] than when preceded by [r]. Second, there was an effect due to liquids having been produced before [d] or [g]: Many more "g" percepts occurred when the preceding liquid had been originally produced before [g] than when it had been produced before [d]; this effect was much stronger for [r] than for [l].

The finding that [l] led to more "g" percepts than [r] is remarkably like a finding observed in studies of the influence of preceding fricatives on stop perception (Mann & Repp, in press-a): [l], which has a more forward place of articulation than [r], leads to relatively more velar stop responses, just as does [s], which has a more forward place of articulation than [ʃ]. The fact that [s] leads to more velar responses than [ʃ] has been attributed to the fact that subjects are, in some sense, aware that stops that follow [s] can receive a relatively more forward place of articulation than those that follow [ʃ]. Perhaps the contrasting effects of [l] and [r] could be similarly explained. Certainly this contrast cannot reasonably be explained in terms of the relative frequencies of various liquid-stop clusters in the English language, especially since the effect operates across a syllable boundary. On the other hand, the present experiment does not eliminate the possibility that the results are due to some auditory interaction involving VC offset and CV onset spectra. For example: the contrasting effects of [l] and [r] could conceivably be the consequence of some form of auditory contrast between the concentration of energy in the F3 region at the end of the preceding VC and that in the F3 region at the beginning of the following CV. Perhaps, the relatively higher F3 offset frequency in [l] led to the perception of a lower F3 onset frequency in the following CV syllable, and thus to more "g" percepts. Nevertheless, the conjecture outlined in the introduction also remains plausible; namely, that stops that follow [l] were more often

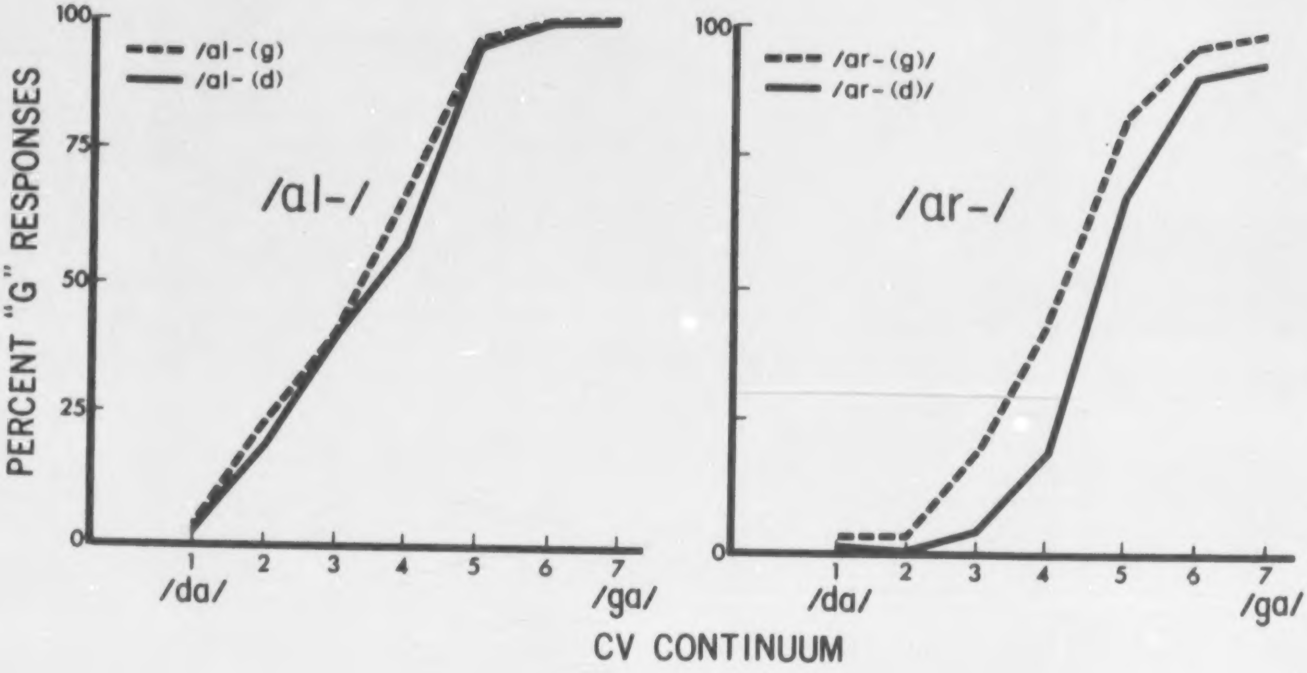


Figure 3. Percentage of "g" responses given to CV stimuli as a function of whether the preceding liquid had originally been produced before [d] or [g].

perceived as "g" because stops that follow [l] tend to be produced with a relatively more forward place of articulation than those that follow [ar].

To gain some support for this contention, we turn to spectrographic measurements of the natural CV syllables from which the test materials were constructed. (See the Appendix for a discussion of the method employed.) Average formant transitions for these syllables are shown in Figure 4, with values for [da] and [ga] represented separately. Comparison of the transitions for stops preceded by [l] (dashed line) with those for stops preceded by [r] (solid line) reveals that stops that followed [l] had greater separation between the onset values of F_2 and F_3. Since velar stops typically show a greater convergence of the onset values for these two formants than alveolar ones, this finding accords with the view that stops that follow [l] can receive a relatively more forward place of occlusion. The extent to which such fronting is typical of all speakers remains an open question. For the moment, however, it is sufficient to note that the present peceptual context effect was obtained with the voice of a speaker who tended to front stops after [l]. Thus, a plausible explanation of the effect of liquid category is that it reflects perceptual compensation for left-to-right, or perseverative, coarticulation in the production of liquid-stop sequences.

The effect due to [al] and [ar] having been produced before [d] or [g] likewise may derive from a coarticulatory influence--but from one that is right-to-left, or anticipatory, in nature. This second effect is also different from the first in that it is a direct consequence of coarticulatory-induced variation in the signal rather than a perceptual compensation for such variation. Thus it is analogous to the finding (Repp & Mann, in press) that, when synthetic stimuli from a [da]-[ga] continuum are preceded by fricative noises excised from naturally-produced fricative-stop sequences, they tend to be perceived as the stop that originally followed the fricative. For fricatives, however, it has further been shown that the acoustic consequence of coarticulation with a following stop is an observable change in noise spectrum. The implication, then, is that when [al] or [ar] preceded velar or alveolar stops, they may have contained cues to the following stop because stop production systematically influenced some aspect of their acoustic structure. The fact that such systematic influences were indeed present can be seen in Table 2, where the average formant offset frequencies are given for [al] and [ar] as a function of whether they preceded [da] or [ga]. (The method used in obtaining these measurements is described in the Appendix.) For both [al] and [ar], offset spectrum was considerably influenced by the place of the following stop. Indeed, the following stop had a relatively greater influence on [ar], which is consistent with the perceptual results obtained with these stimuli. The fact that listeners are able to make correct use of such influences as cues to stop perception attests to the view that speech perception must somehow operate with tacit reference to the dynamics of speech production and its acoustic consequences. How else can we explain the fact that such a multiplicity of cues seem capable of influencing stop consonant perception? The commonality between those cues is neither their acoustic structure nor their location in time, but rather that they reflect one and the same "articulatory act" (Repp et al., 1978).

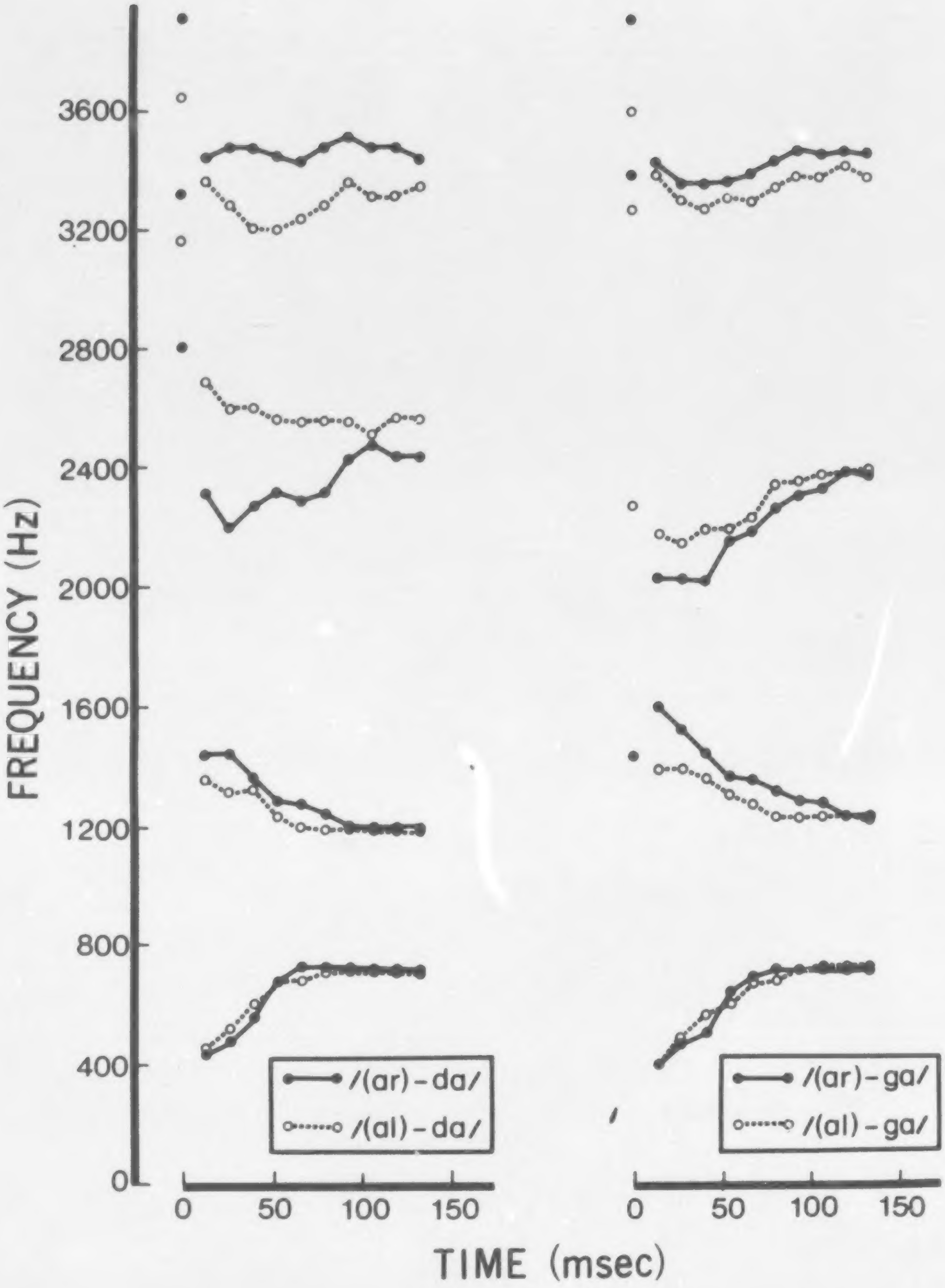


Figure 4. Average formant values for the first 145 msec of natural [da] and [ga], plotted separately for tokens produced after [al] and [ar].

Table 2

Average Formant Offset Frequencies in Naturally-Produced VC Syllables
(Standard deviations in parentheses)

	F_1	F_2	F_3	F_4
[ar-(da)]	400(67)	1473(17)	1680(143)	2727(89)
[ar-(ga)]	407(30)	1306(49)	1786(106)	3453(218)
[al-(da)]	447(141)	927(40)	2773(41)	3553(119)
[al-(ga)]	420(49)	1020(79)	2649(39)	3573(200)

In summary, the high degree of consistency between the present perceptual findings and the dynamics of speech production is reminiscent of that seen in several previous studies of contextual influences on stop consonant perception. Clearly, the conclusion to be drawn from this consistency is that the observed influences of liquid context reflect listeners' sensitivity to the coarticulatory influences involved in the production of liquid-stop sequences. There are two aspects of this sensitivity that are particularly relevant to our understanding of the type of mechanisms that must be accomplishing human speech perception: First, that perception takes into account coarticulatory influences in both directions, that is, from left-to-right and right-to-left; and second, that it can operate across a well-defined syllable boundary. These results, which cannot easily be explained by models of speech perception that postulate either phoneme- or syllable-sized templates, accord with the view that speech perception is an active process guided by some tacit knowledge of articulatory dynamics.

REFERENCES

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. Stop consonant perception: Release bursts and formant transitions as functionally equivalent, context dependent cues. *Perception & Psychophysics*, 1977, *22*, 109-122.

Clark, H. H. The language-as-a-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335-359.

Gay, T. Cinefluorographic and electromyographic studies of articulatory organization. *Haskins Laboratories Status Report on Speech Research*, 1977, *SR-50*, 77-92.

Kuhn, G. M. Stop consonant place perception with single-formant stimuli: Evidence for the role of the front-cavity resonance. *Journal of the Acoustical Society of America*, 1977, *65*, 774-788.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. The role of selected variables in the perception of the unvoiced stop consonants. American Journal of Psychology, 1952, 65, 497-516.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. Journal of the Acoustical Society of America, in press. (a)

Mann, V. A., & Repp, B. H. Influence of vocalic context on the [ʃ]-[s] distinction. Perception & Psychophysics, in press. (b)

Öhman, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, 1966, 39, 151-168.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 621-637.

Repp, B. H., & Mann, V. A. Perceptual assessment of fricative-stop coarticulation. Journal of the Acoustical Society of America, in press.

APPENDIX

In measuring the formant frequencies of the naturally produced syllables, I relied on spectral cross-sections generated by a Federal Scientific UA-6A spectrum analyzer and displayed as point plots on a Hewlett-Packard 1300 Oscilloscope, together with a computer-generated spectrogram and wave-form display. All spectral information was smoothed and pre-emphasized. The cross-sections were derived from 25.6-msec windows in 12.8-msec steps. The precise location of the first window could not be controlled; thus, the first section of each syllable usually included some of the silence preceding the utterance, and spectral peaks usually were not evident until the second section. The location of peaks for the first four formants was estimated visually, the maximum resolution being 40 Hz.

Two portions of each disyllable were of particular interest: the offset of the VC syllable, and the transitions in the CV syllable. For each portion, I determined formant values that were subsequently averaged across the three tokens of each disyllable in each of the two stress patterns. Spurious peaks that were not common to all six tokens were omitted. Table 2 gives the average formant values for the last cross-section of the VC syllable that contained peaks for each of the first four formants. Figure 4 shows the formant values for the initial 12 sections of the CV syllable, starting with the first section with measurable spectral energy.

PERCEPTUAL ASSESSMENT OF FRICATIVE-STOP COARTICULATION*

Bruno H. Repp and Virginia A. Mann+

Abstract. The perceptual dependence of stop consonants on preceding fricatives (Mann and Repp, in press) was further investigated in two experiments employing both natural and synthetic speech. These experiments consistently replicated our original finding that listeners report more velar stops following [s]. In addition, our data confirmed earlier reports that natural fricative noises (excerpted from utterances of [stɑ], [skɑ], [ʃtɑ], and [ʃkɑ]) contain cues to the following stop consonants; this was revealed in subjects' identifications of stops from isolated fricative noises and from stimuli consisting of these noises followed by synthetic CV portions drawn from a [tɑ]-[kɑ] continuum. However, these cues in the noise portion could not account for the contextual effect of fricative identity ([ʃ] vs. [s]) on stop perception (more "k" responses following [s]). Rather, this effect seems to be related to a coarticulatory influence of a preceding fricative on stop production: Subjects' responses to excised natural CV portions (with bursts and aspiration removed) were biased towards a relatively more forward place of stop articulation when the CVs had originally been preceded by [s]; and the identification of a preceding ambiguous fricative was biased in the direction of the original fricative context in which a given CV portion had been produced. These findings support an articulatory explanation for the effect of preceding fricatives on stop consonant perception.

INTRODUCTION

In a recent paper (Mann & Repp, in press), we described a perceptual dependency of stop consonants on preceding fricatives: a stop ambiguous between [t] and [k] was more likely to be labeled "k" when preceded by [s] than when preceded by [ʃ] or by no fricative at all. This perceptual context effect was demonstrated in a series of experiments with synthetic speech. The present experiments employed both natural and synthetic speech to investigate further the possible origins of this effect.

---

*To appear in the *Journal of the Acoustical Society of America* in a revised form.

+Also at Bryn Mawr College.

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories, and by NICHD Postdoctoral Fellowship HD05677 to Virginia A. Mann. Portions of the results were reported at the 99th Meeting of the Acoustical Society of America in Atlanta, April 1980. We thank Alvin Liberman, Michael Studdert-Kennedy, and Winifred Strange for helpful comments.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

We proposed in our earlier paper that the influence of fricative context on stop perception reflects listeners' perceptual compensation for a coarticulatory influence of fricatives on following stop consonants; an influence which results in a relative forward shift of velar and/or alveolar place of stop occlusion following [s]. Of course, the most direct ways of confirming the existence of such a coarticulatory effect would be to observe ongoing articulation and to measure its consequences in the acoustic signal. We are engaged in such efforts and hope to report their outcome in a separate paper. The present experiments, however, took a more indirect approach. Their purpose was to provide perceptual evidence for coarticulation by excerpting portions from natural utterances and examining how listeners identify them, both when presented in isolation and when recombined with (more or less ambiguous) synthetic stimulus portions. Such perceptual assessment of coarticulation, while it cannot replace direct articulatory and acoustic measurements, has the special advantage of revealing whether a given coarticulatory effect has any perceptual significance.

Several previous studies have attempted to assess coarticulation by excerpting acoustically defined segments from natural utterances and presenting them to listeners for identification. For example, Fant, Liljencrants, Malac, and Borovickova (1970) and Lehiste and Shockey (1972) used this method to find evidence for effects of different initial vowels on the opening transitions (and of different final vowels on the closing transitions) of stops in VCV utterances; it was used by Benguerel and Adelman (1975) and by Yeni-Komshian and Soli (1979) to find perceptually significant traces of vowel quality in preceding consonants; and by Ali, Gallagher, Goldstein, and Daniloff (1971) to determine the detectability of vowel nasality due to following nasal consonants. This technique has serious drawbacks, however. When listeners are required to identify phonetic segments whose primary cues have been deleted from the speech signal, the task becomes one of inference or guessing rather than perception. On the other hand, when listeners merely report the phonetic segments they actually perceive, performance is often too accurate to be sensitive to small variations in signal parameters.

We have used the "method of isolation" with some success in the present studies (Exps. 1B and 2A); however, we have relied, in addition, on a second, novel method which we find especially attractive--the "method of substitution" (Exps. 1B, 2C, and 2D). Instead of omitting a portion of the signal, we replace it with a phonetically ambiguous, synthetic stimulus of similar overall structure. We then test for the presence of perceptually significant coarticulatory traces in the remaining natural signal portion by gauging their power to bias perception of the ambiguous synthetic stimulus towards the phonetic category corresponding to the replaced segment. Thus, the synthetic substitute may serve as an indicator of coarticulatory effects, and useful results may be obtained where the method of isolation would yield only chance-level guessing or near-perfect identification.[1]

Below we report two experiments. The first employed natural fricative noises that were excerpted from fricative-stop-vowel (FCV) utterances. By presenting these noises in isolation and in conjunction with synthetic CV portions, we examined the role of coarticulatory cues to stop identity in the fricative noise portion. The second experiment employed natural CV portions from the same FCV utterances. By presenting these stimuli in isolation and in conjunction with synthetic fricative noises, we endeavored to determine whether CV portions contain coarticulatory traces of the fricative that originally preceded them. Our experiments provide clear perceptual evidence that such traces exist, thus corroborating our hypothesis (Mann & Repp, in press) that the perceptual influence of preceding fricatives on stop consonant perception has a basis in coarticulation.

EXPERIMENT 1

Experiment 1 had three conditions (A, B, C). Those methodological aspects common to all three are described below; specific features are described later under individual headings.

General Method

Subjects. Ten subjects participated. They included seven paid volunteers (some of whom had taken part in earlier experiments employing similar stimuli), a research assistant, and the two authors. Since experience did not seem to influence the basic pattern of results, the data were pooled across subjects in this and subsequent experiments.

Stimuli. A male, phonetically trained, native speaker of American English spoke the utterances [sɑ], [ʃɑ], [stɑ], [skɑ], [ʃtɑ], [ʃkɑ] repeatedly in random order as part of a list containing a number of other utterances. The recordings were made in a soundproof booth using a Shure dynamic microphone and a calibrated Ampex AG-500 tape recorder. Subsequently, the utterances were digitized at 10 kHz and stored in separate files using the Haskins Laboratories Pulse Code Modulation (PCM) system. Three good tokens of each of the six utterances were selected for use in the experiments. The fricative noise was excerpted from each stimulus and stored separately. Acoustic parameters of these noises are given in the Appendix.

In Conditions A and C, some of the natural fricative noises were combined with digitized synthetic CV portions drawn from a [tɑ]-[kɑ] continuum that had been created on the OVE IIIc synthesizer at Haskins Laboratories. There were seven CV stimuli, distinguished only by the onset frequency of the third formant (F_3) which decreased from 3222 Hz for the most [tɑ]-like stimulus to 1902 Hz for the most [kɑ]-like stimulus in steps of approximately 215 Hz (plus or minus up to 10 Hz). All stimuli had 50-msec stepwise-linear formant transitions (F_1: from 285 to 771 Hz; F_2: from 1770 to 1233 Hz; F_3: to 2520 Hz) followed by 200 msec of steady-state resonances, a linearly falling fundamental frequency (110 to 80 Hz), and a flat amplitude contour with a 50-msec ramp at onset and a 30-msec ramp at offset. These stimuli were perceived as /dɑ/ or /gɑ/ in isolation but as /tɑ/ or /kɑ/ when preceded by a fricative noise, due to the phonotactic principles of English.

Procedure. The subjects listened to the stimulus tapes (described below) in a quiet room at a comfortable intensity, using an Ampex AG-500 tape recorder and Telephonics TDH-39 earphones. The conditions were presented in a single session in fixed order (A, C, B), separated by brief rest periods.

Condition A: Replication of basic context effect

The purpose was to replicate the basic finding that listeners are biased to hear "k" rather than "t" in the context of a preceding [s], as compared with a preceding [ʃ] or a null context. To avoid the problems inherent in synthesizing appropriate fricative noises (Mann & Repp, in press), we used natural fricative noises in conjunction with a synthetic [tɑ]-[kɑ] continuum.

Method. Listeners first heard a sequence of isolated CV syllables (the seven stimuli from the [tɑ]-[kɑ] continuum ten times in random order) that they identified as beginning with "d" or "g". Subsequently, they listened to the same syllables preceded by a fricative noise plus a 75-msec silent interval. The noises were those excerpted from [ʃɑ] and [sɑ], and there were three tokens of each. As there were six physically different noises, there were 42 different stimulus combinations that were presented five times in random order. The subjects identified both the fricative ("sh" or "s") and the stop ("t" or "k").

Results and discussion. Figure 1 shows the results. Because of the rather wide spacing of the stimuli on the synthetic [tɑ]-[kɑ] continuum, listeners' category boundaries were quite sharp, so that the present test of effects of fricative context was conservative. Of the seven CV syllables, only stimulus 4 was ambiguous in isolation, and it was the only one whose perception was affected by a preceding fricative. However, that effect was exactly as predicted: a preceding [ʃ] had no effect relative to the isolated-CV baseline whereas a preceding [s] lowered the percentage of "t" responses. This small effect was sufficiently consistent across subjects to be highly significant in a standard repeated-measurements analysis, $F(1,9) = 20.4$, $p < .005$, and it also reached significance when the variation between fricative noise tokens was taken as the error estimate, $F(1,4) = 11.3$, $p < .05$.[2]

Thus, we successfully replicated the basic effect of a preceding [s] on stop consonant perception. By replicating the effect with natural fricative noises, we have eliminated any doubts deriving from our earlier use of synthetic noise stimuli. However, the possibility still exists that the natural [s] and [ʃ] noises were not equally neutral as potential cues to place of articulation of a following stop. The next experiment addressed this point.

Condition B: Identification of stops from F(CV) portions

In part, this condition examined how accurately listeners can identify alveolar and velar stop consonants upon hearing fricative noises excerpted from FCV utterances. That cues to stop place of articulation are contained in fricative noises that precede a stop closure has been reported by several researchers (Uldall, 1964; Malécot & Chermak, 1966; Schwartz, 1967; Bailey & Summerfield, 1980). These cues consist of spectral shifts ("transitions") due to progressive narrowing of the vocal tract towards the stop occlusion (see our Appendix). Malécot and Chermak (1966) and Schwartz (1967) have shown that listeners can identify stops fairly accurately from isolated fricative noises containing appropriate spectral shifts. However, the stop most accurately identified is [p], which was not included in our materials. Earlier studies suggest that [t] and [k] are more difficult to identify from fricative-noise

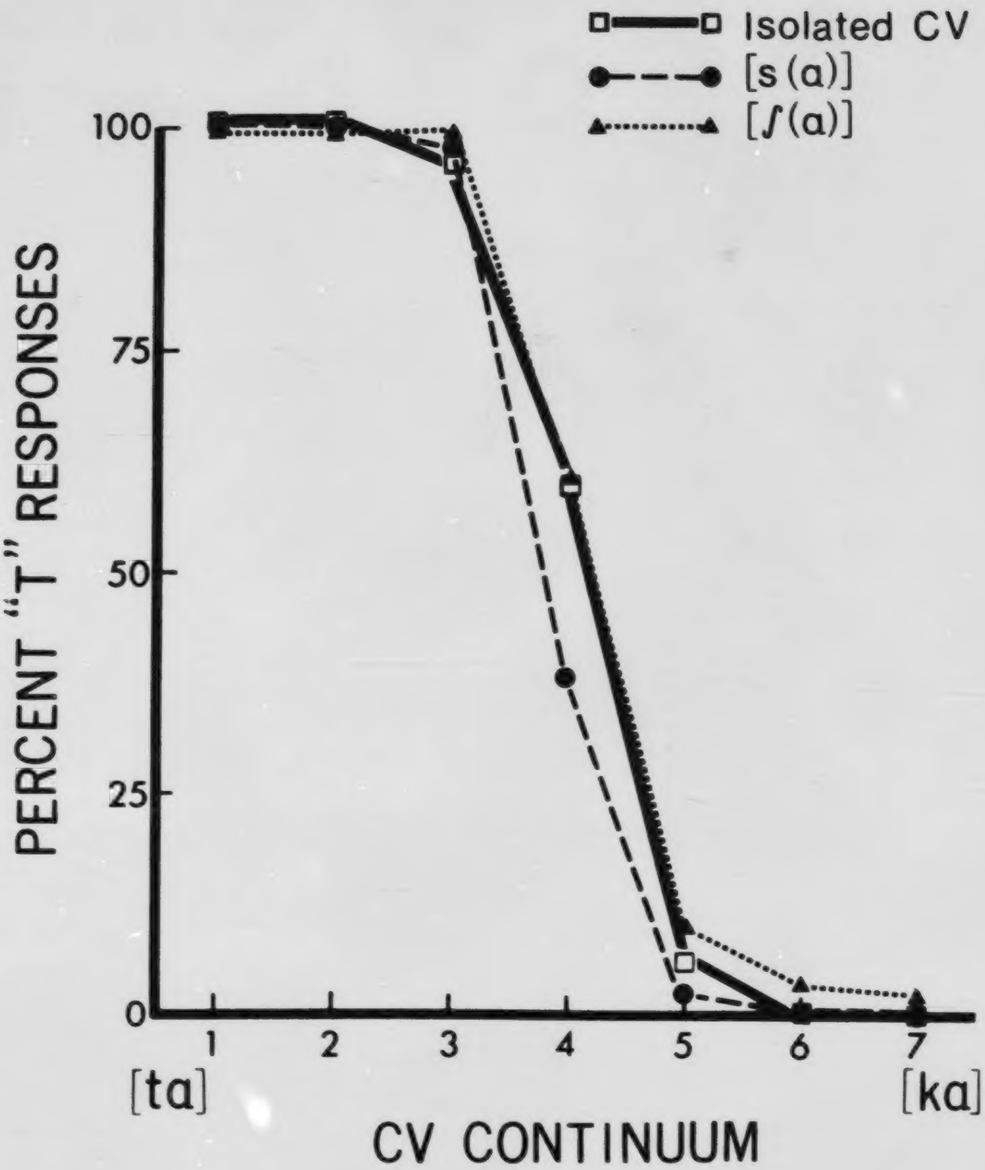


Figure 1. Effects of preceding [ʃ] or [s] (without cues to stop manner) on stop consonant perception (Exp. 1A).

transitions alone. Since we were concerned about the potential role of these cues in the influence of preceding fricatives on stop perception, it was important to determine just how salient these cues were.

In addition to the noises excerpted from FCV utterances, we included the noises used in Condition A, which derived from FV utterances. We wondered whether listeners' forced-choice stop responses to these latter noises would exhibit a bias towards "k" following [s]. Such a bias would suggest that these noises were not equally neutral as potential cues to place of stop occlusion; or, considering the fact that these noises really did not contain any such cues (according to our own perception and acoustic analysis--see the Appendix), a response bias contingent on fricative identity would be implicated.

Method. The fricative noises were excerpted from natural [ʃɑ], [sɑ], [ʃtɑ], [stɑ], [ʃkɑ], [skɑ]. As there were three different tokens of each noise, there were 18 stimuli altogether that were presented five times in random order. The subjects' task was to identify the fricative as "sh" or "s" and, in addition, to report (or guess) whether that fricative had been originally followed by "t" or "k". The subjects were told that all noises had been excerpted from FCV utterances; they were not informed about the fact that some derived from FV utterances.

Table 1

Percentages of "t" and "k" responses to isolated fricative noises

Stimulus	Response	
	"t"	"k"
[ʃ(tɑ)]	91.3	8.7
[ʃ(kɑ)]	30.7	69.3
[s(tɑ)]	94.7	5.3
[s(kɑ)]	12.7	87.3
[ʃ(ɑ)]	40.7	59.3
[s(ɑ)]	54.0	46.0

Results and discussion. The results are shown in Table 1. Considering first only the noises derived from FCV utterances, it is clear that the subjects could identify the stop consonants quite well, being correct on 86 percent of the trials. They were more accurate in identifying [tɑ] than [kɑ], $F(1,9) = 11.6$, $p < .01$. They were also somewhat more accurate with stops following [s] rather than [ʃ], $F(1,9) = 8.4$, $p < .05$, particularly where "k" responses were concerned. Both effects were equally significant with token variability as the error estimate. The second effect could be taken as a bias to respond "k" in conjunction with "s". However, the statistical interaction that would have supported such a bias was not significant. Moreover, the responses to the noises deriving from FV utterances did not suggest such a bias: "k" responses were actually more frequent in conjunction with [ʃ] than with [s], although the difference did not reach significance. Furthermore, we note that "k" responses to FV noises were slightly more frequent than "t" responses. This indicates that the better identification of [tɑ] than [kɑ] in FCV noises was due to the nature of the acoustic information--possibly the absence of [k]-release bursts (cf. Malécot & Chermak, 1966)--and not to a simple response preference for "t".[3]

Thus, we find no evidence of a bias to respond "k" in conjunction with "s" when isolated fricative noises are presented. Apparently, the presence of a full FCV stimulus is necessary to evoke that tendency; therefore, it seems unlikely that we are dealing with a response bias contingent on fricative identity.

Condition C: Dissociating two effects of preceding fricatives on stop perception

As a further test of the role of cues to place of stop occlusion in the fricative noise, we juxtaposed fricative noise transitions with CV formant transitions, both of which may serve as cues to place of stop occlusion in FCV stimuli. When conflicting vocalic formant transitions are juxtaposed (VC-CV), the CV transitions generally dominate perception; or, if the closure interval is sufficiently long (70 msec or more), two different stop consonants are heard in sequence (Repp, 1978; Dorman, Raphael, & Liberman, 1979). By analogy, we expected the noise transitions to be less salient as cues to stop place of articulation than the CV transitions; the question was whether the noise transitions would have any effect whatsoever. At the silent interval used here (75 msec), we did not notice any tendency to hear two different stops ([stkɑ] or the like).

Whether or not listeners assigned any perceptual weight to the fricative noise cues, we expected to find the basic contextual effect of fricative identity on stop perception (more "k" responses following [s]). By aiming at replicating the context effect using natural fricative noises containing appropriate cues to stop articulation, the study effectively avoided the problem of having to decide whether [ʃ] and [s] noises without such cues are equally "neutral" (cf. Condition A). Instead, fricative identity and noise transitions were treated as independent variables in a 2 x 2 factorial design.

Method. The seven stimuli from the [tɑ]-[kɑ] continuum were preceded by [ʃ] or [s] noises excerpted from natural [ʃtɑ], [ʃkɑ], [stɑ], [skɑ], with 75 msec of silence in between. As there were three physically different noises from each context--12 noises in all--there were 84 stimulus combinations that were presented five times in random order. The subjects identified both the fricatives ("sh" or "s") and the stops ("t" or "k").

Results and discussion. Figure 2 shows that, despite the relatively sharp category boundary on the [tɑ]-[kɑ] continuum, there were clear effects of the fricative noise on stop identification. First of all, noise transitions did influence stop identification: there were more "t" responses with

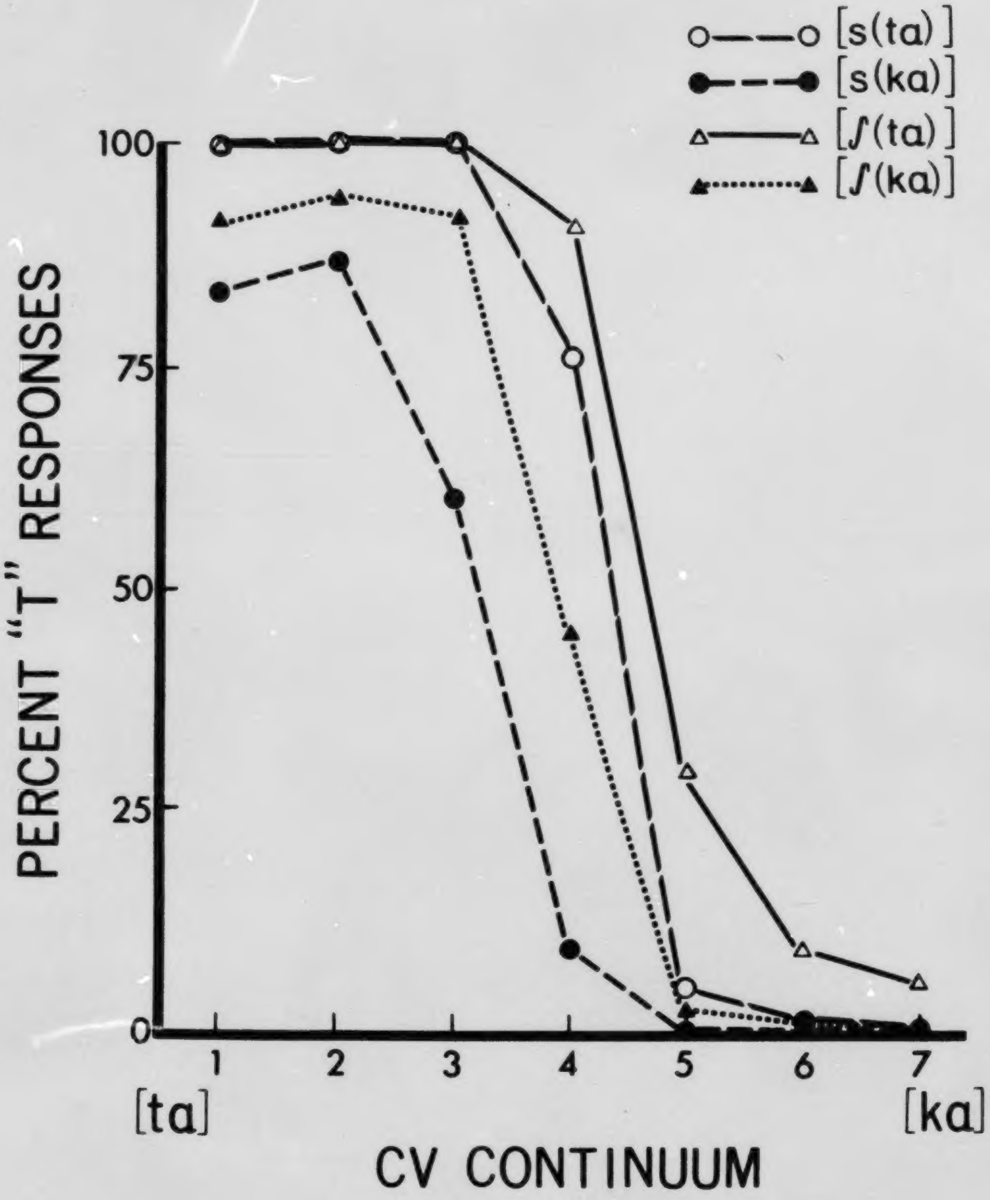


Figure 2. Effects of preceding [ʃ] or [s] (with cues to stop manner and place of articulation) on stop consonant perception (Exp. 1C).

transitions deriving from [t] than with transitions deriving from [k], $F(1,9) = 26.6$, $p < .0005$. As predicted, however, the CV transitions were the stronger cue to stop place of articulation, for the noise transitions had relatively small effects when the CV transitions were unambiguous. Second, the basic context effect was replicated: there were more "t" responses following [ʃ] than following [s], $F(1,9) = 31.5$, $p < .0005$. Finally, the two effects did not interact statistically, $F(1,9) = 4.7$, $p > .05$, and thus appeared to be independent. The same results were obtained in an analysis by tokens, since token variability was generally small.[4]

Thus, the present data show that the basic context effect of a preceding fricative on stop perception is obtained even when there are cues to place of stop occlusion in the fricative noise portion. This reinforces our earlier conclusion (Mann & Repp, in press) that there is a context effect due to the fricative *per se*, which is independent of noise properties that directly reflect stop production.

EXPERIMENT 2

The results of Experiment 1 suggest that the contrasting effect of preceding [ʃ] and [s] on stop perception reflects neither a simple response bias nor an effect of cues to stop place of articulation contained in the fricative noise. By ruling out these alternatives, we have gained indirect support for our hypothesis that the effect derives from perceptual compensation for a coarticulatory influence of a preceding fricative on stop consonant production. In our second experiment, which comprised four conditions, we attempted to obtain direct evidence for such a coarticulatory dependency by examining in several different ways how listeners respond to natural CV portions that had been originally produced in the context of either [ʃ-] or [s-].

General Method

Subjects. Twelve subjects participated in Conditions A, B, and D, which were run in a single session in a fixed order (B, D, A). There were nine paid volunteers, two of whom had been subjects in Experiment 1, plus a research assistant and the two authors, all of whom had been subjects in the earlier experiment. These last three subjects participated in two identical sessions whose results were averaged before they were combined with the results of the other subjects, who participated only in a single session. Condition C was conducted at a later time and used a partially different group of ten subjects (seven new paid volunteers, the research assistant, and the two authors).

Stimuli. The same natural utterances of [ʃtɑ], [ʃkɑ], [stɑ], and [skɑ] that had supplied the fricative noises of Experiment 1 also provided the CV portions for the present experiments. There were three physically different tokens of each CV stimulus, and each was employed in two versions, one including the release burst and one without the burst. The stimuli with bursts consisted of the total signal portion following the silent closure interval in the source utterances. The burstless stimuli were obtained by deleting all energy preceding the first clear pitch pulse; the deleted portion usually included a small amount of aspiration following the release burst. All in all, there were 24 distinct CV portions. Details of their acoustic structure are reported in the Appendix.

In Conditions C and D, these CV portions were preceded by synthetic fricative noises from a nine-member [ʃ]-[s] continuum; in Condition B, just the endpoints of that continuum were used. The fricative noises were distinguished by the center frequencies of two poles generated by the fricative circuit of the OVE IIIc synthesizer. (No zero was specified.) These frequencies are listed in Table 2; they increased in roughly equal steps from stimulus 1 ([ʃ]-like) to stimulus 9 ([s]-like). All noises were 200 msec in duration and had approximately equal amplitudes, with a triangular amplitude contour that peaked after 150 msec. They were digitized at 10 kHz.

Table 2

Pole frequencies of fricative noises (Hz)[a]

Stimulus	Pole 1	Pole 2
[ʃ] 1	1957	3803
2	2197	3915
3	2466	4148
4	2690	4269
5	2933	4394
6	3199	4655
7	3389	4792
8	3591	4932
[s] 9	3917	5077

[a] The values given are synthesizer input parameters. Measurements of the acoustic output suggested that the actual pole center frequencies were about 5 percent lower. Some irregularities in step size were caused by our use of prespecified frequency values in conjunction with the limited frequency resolution of the synthesizer. Any effect these irregularities might have had on our results in Experiments 2C and 2D worked in favor of the null hypothesis.

Condition A: Identification of stops from (F)CV portions

This condition provided the most direct perceptual test for the existence of coarticulatory variations in the production of stops following [ʃ] and [s]. In this study, which used the "method of isolation," the subjects' task was to identify the initial (stop) consonants in isolated CV portions, with and without bursts. To the extent that any confusions would occur along the place dimension, we expected these errors to reflect any coarticulatory variation in the CV formant transitions (and perhaps in the release burst) introduced by the original fricative context. Specifically, since coarticulation is generally assimilative, a stop following [s] might exhibit transitions reflecting a more forward place of articulation than a stop following [ʃ], because [s] has a more forward place of articulation than [ʃ]. Therefore, if such coarticulatory effects exist, we expected errors in stop identification to be biased towards a forward place of articulation when the CV portion had originally been preceded by [s]. It was considered possible that this effect, if obtained, would be more pronounced for (intended) [k] than for [t], since velar place of articulation might have more freedom to shift than alveolar place of articulation (as evidenced by the existence of two major allophones of velar stops in English). Also, judging from our earlier perceptual results and from our introspections on fricative-stop articulation, coarticulatory shifts in stop place of articulation should be primarily due to [s]. This implies that stops originally preceded by [ʃ] should be more accurately identified than those originally preceded by [s].

Method. The 24 CV portions (two intended stops, two original fricative contexts, three tokens, with and without burst) were presented five times in random order. The listeners had to identify the initial consonant by forced choice between four alternatives: "b", "th", "d", "g". It was explained that "th" represented the initial sound in that; this fricative, whose place of articulation is--roughly speaking--intermediate between "b" and "d", is easily perceived in the absence of any fricative noise and was in fact a frequent response choice.

Table 3

Identification of stops in isolated CV portions

	Without burst				With burst			
Stimulus	Response (percent)							
	"b"	"th"	"d"	"g"	"b"	"th"	"d"	"g"
[(s)tɑ]	28.0	52.2	17.2	2.5	---	21.1	78.9	---
[(ʃ)tɑ]	5.6	44.7	43.3	6.4	---	8.6	86.6	4.7
[(s)kɑ]	13.6	15.3	21.4	49.7	---	---	---	100.0
[(ʃ)kɑ]	2.2	2.8	13.3	81.7	---	---	---	100.0

Results and discussion. The listeners' responses are summarized in Table 3. First of all, it is immediately evident that stimuli with bursts were much more accurately identified than burstless stimuli. When bursts were present, misidentifications occurred only with [tɑ], and they were primarily "th" responses. These responses, however, were more frequent to [(s)tɑ] than to [(ʃ)tɑ], which is in accord with our hypothesis that [(s)tɑ] has a more forward place of articulation than [(ʃ)tɑ].

This hypothesis is further supported by the pattern of responses to burstless stimuli, which were much less accurately identified. First, we see that [kɑ] was more often "correctly" identified as "g" than [tɑ] as "d", $F(1,11) = 10.8$, $p < .01$--an unexpected result that was apparently due to the large number of "th" responses to [tɑ] stimuli.[5] Second, more "errors" occurred in response to [(s)-] stimuli than to [(ʃ)-] stimuli, $F(1,11) = 19.4$, $p < .002$. Since virtually all errors were in the direction of a more forward place of articulation (except for the rare "g" responses to [tɑ]), the result implies that [(s)-] stimuli had a more forward place of production than [(ʃ)-] stimuli, as predicted.

There were some marked differences between individual stimulus tokens. In particular, one of the three burstless [(s)kɑ] tokens evoked the response pattern characteristic of the [(ʃ)kɑ] tokens. This indicates a fair amount of articulatory variability from utterance to utterance. However, with token variance as the error term, the differences just reported were still significant at the $p < .05$ level.

These results confirm our hypothesis of a forward shift in place of stop articulation following [s], and, moreover, are in accord with our perceptual results in suggesting that the shift is, indeed, primarily due to [s]. We cannot tell from these results whether the release bursts conveyed any information about these articulatory shifts since, in most cases, the presence of a burst seemed to be sufficient for correct identification; therefore, whatever spectral variations occurred in the burst portion were not revealed in listeners' responses. However, the vocalic formant transitions must have varied with the preceding fricative in the manner predicted (see the Appendix), and this variation was, moreover, perceptually significant. Thus, we now have support for an articulatory effect that parallels the perceptual context effect observed in our earlier studies.

Condition B: Identification of stops in F+(F)CV stimuli

In this condition, the CV stimuli of Condition A were presented in the context of an actual preceding [ʃ] or [s]. Thus, in addition to recreating (approximately) the context in which the stops had been originally produced, we had the opportunity to observe any effect of preceding synthetic fricative noises on the perception of stops cued by natural CV portions.

Method. The 24 natural CV portions were preceded by either a [ʃ]-noise or a [s]-noise, the endpoint stimuli of a synthetic noise continuum (see Table 2), plus a 75-msec silent gap. The resulting 48 stimuli were presented five times in random order. The subjects' task was to identify the fricative as either "sh" or "s", and the stop as either "p", "t", or "k". Note that, in the context of a preceding fricative, the stops were now to be given voiceless category labels, in conformity with the phonology of English. In contrast to Condition A, "th" responses did not seem appropriate here, as [sθ] and [ʃθ] clusters are extremely uncommon and not readily perceived.

Results and discussion. The results are displayed in Table 4, separately for stimuli preceded by synthetic [ʃ] and stimuli preceded by synthetic [s]. The fricatives were generally correctly identified (2.7 percent errors). Without the "th" response category, the stops in stimuli with bursts were now identified with nearly perfect accuracy, and burstless [tɑ] was now identified more accurately than burstless [kɑ], as originally predicted, at least when preceded by [ʃ]. (However, see Footnote 5.) Otherwise, the responses to burstless stimuli replicated the pattern found in Experiment 4: The stops in [(ʃ)-] stimuli were identified more accurately than the stops in [(s)-] stimuli, and confusions for [(s)-] stimuli tended more towards a forward place of articulation than confusions for [(ʃ)-] stimuli, $F(1,11) = 7.2$, $p < .05$, this effect being most pronounced for [k]. In addition, there was a clear effect of the preceding synthetic fricative noise: "t" responses were more frequent after [ʃ], while "k" responses were more frequent after [s], $F(1,11) = 34.7$, $p < .001$. Thus, the present experiment replicated both the coarticulatory effect (due to the excerpted fricative) and the corresponding perceptual effect (due to the substituted fricative) on stops in a single design. The marked token differences observed in Experiment 4 were also replicated; however, all statistical effects held up when token variance was taken as the error estimate.

Table 4

Stop identification in natural CV portions preceded by synthetic [ʃ] or [s]

Stimulus	Without burst			With burst		
	Response (percent)					
	"p"	"t"	"k"	"p"	"t"	"k"
[ʃ]+[(s)tɑ]	10.5	83.8	5.6	0.3	99.7	---
[ʃ]+[(ʃ)tɑ]	3.0	91.7	5.3	---	98.3	1.7
[ʃ]+[(s)kɑ]	4.7	66.4	28.9	---	1.7	98.3
[ʃ]+[(ʃ)kɑ]	1.1	51.1	47.8	---	4.4	95.6
[s]+[(s)tɑ]	10.5	74.4	15.0	0.3	99.1	0.6
[s]+[(ʃ)tɑ]	5.0	79.4	15.6	---	96.4	3.6
[s]+[(s)kɑ]	3.0	31.7	65.3	---	---	100.0
[s]+[(ʃ)kɑ]	---	20.0	80.0	---	---	100.0

Condition C: Fricative identification in F+(F)CV stimuli

In this study, we employed the "method of substitution" to see whether the coarticulatory traces of the preceding fricatives in the natural CV portions would bias the perception of ambiguous synthetic fricative noises in the direction of the original fricative. Thus, this experiment was analogous to Experiment 1C, which showed that cues contained in natural fricative noises that had been excised from FCV utterances influenced stop perception when synthetic CV portions were added. There is an important difference, however: The cues to place of stop articulation in the fricative noise of an FCV utterance are quite pronounced and, as we showed in Experiment 1B, generally sufficient to identify the stop from the fricative noise alone. On the other hand, any cues to place of fricative articulation contained in the CV portion are subtle and indirect; our informal observation is that they are not sufficient to identify a missing fricative. Therefore, we expected that any influence of the CV portion on fricative perception would be rather small.

Method. The 24 CV portions were preceded by nine synthetic fricative noises forming an [ʃ]-[s] continuum (Table 2), plus a 75-msec gap. The resulting 216 stimuli were presented four times in random order. The subjects' task was to identify the fricative as "sh" or "s" and the stop as "p", "t", or "k".

Since seven of the ten subjects in Condition C were newly recruited, this part of Experiment 2 also served as a semi-independent replication of the error patterns in stop identification observed in Conditions A and B. In addition, we re-examined a question that received conflicting answers in our earlier studies (Mann & Repp, in press): whether, and in which way, stop identification is influenced by the precise spectral properties of the preceding (steady-state) fricative noise.

Results and discussion. The fricative identification results are shown in Figure 3, separately for stimuli with and without bursts at CV onset. Although the differences between the various identification functions were relatively small, the statistical analysis (conducted on percent "sh" responses averaged over all members of the fricative noise continuum) revealed several reliable effects. First, "sh" responses were more frequent to burstless stimuli than to stimuli containing bursts, $F(1,9) = 12.5$, $p < .01$. Second, "sh" responses were more frequent to stimuli containing [tɑ] than to stimuli containing [kɑ], $F(1,9) = 8.8$, $p < .05$. Third, and most interestingly, "sh" responses were more frequent to stimuli containing [(ʃ)-] CV portions than to stimuli containing [(s)-] CV portions, $F(1,9) = 20.5$, $p < .01$; this was the effect of original fricative context we were looking for. However, there was also a triple interaction, $F(1,9) = 14.0$, $p < .01$. To clarify this interaction, separate analyses were conducted on stimuli with and without bursts.

The separate analyses revealed, for burstless stimuli, only an effect of original fricative context, [(ʃ)-] vs. [(s)-], $F(1,9) = 5.7$, $p < .05$; however, stimuli with bursts showed not only the same effect in more pronounced form, $F(1,9) = 14.5$, $p < .01$, but also an effect of (intended) stop, [tɑ] vs. [kɑ], $F(1,9) = 9.5$, $p < .05$, and an interaction between these two effects, $F(1,9) = 10.3$, $p < .02$. Figure 3 shows that the interaction derives from the effect of original fricative context being larger for [kɑ] than for [tɑ]. Analyses using token variance as the error term yielded the same pattern of results, with somewhat reduced levels of significance; the effect of original fricative context, which was of greatest interest to us, remained significant overall ($p < .01$), and separately for stimuli with bursts ($p < .05$).

These results show that acoustic variations at the onset of the CV portion, induced by the articulation of a preceding fricative noise, are

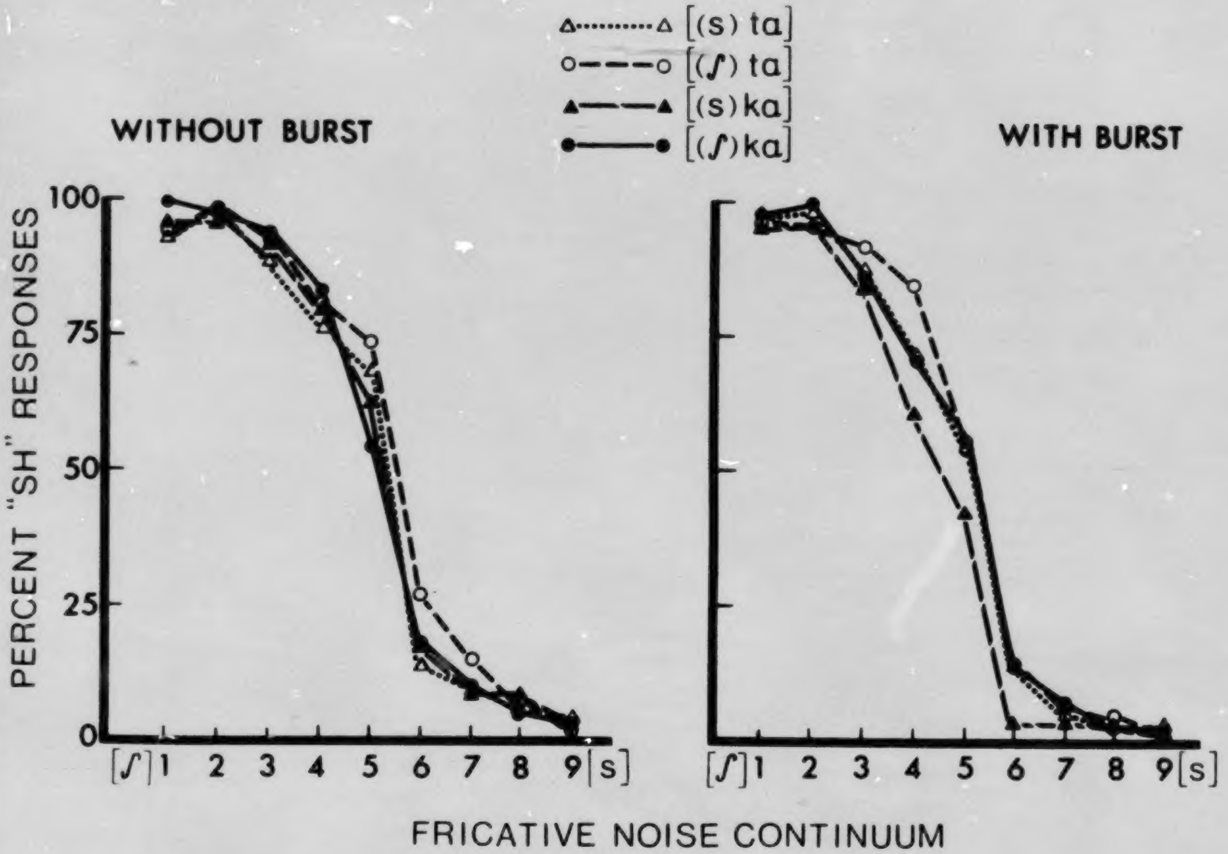


Figure 3. Effects of following CV portions from different fricative contexts on fricative perception (Exp. 2C).

sufficient to create a slight but significant bias towards perception of the original fricative category when an ambiguous noise cue is present. This bias was larger when the CV portion included a burst; thus, the burst may convey part of the coarticulatory information. The finding that "sh" responses were somewhat more frequent with [tɑ] (and "s" responses with [kɑ]) replicates an effect of stop consonant identity on fricative perception that we had observed in one of our earlier studies (Mann & Repp, in press: Exp. 5). The effect mirrors the now-familiar influence of the fricative on stop perception: in both cases, "s" tends to go with "k", and "sh" with "t". That the effect was reliably observed only in stimuli with bursts probably relates to the fact that only these stimuli permitted accurate identification of the intended stops. We have no explanation at present for our finding of an overall increase in "sh" responses in the absence of bursts.

As in Conditions A and B, stop identification was much more accurate when bursts were present: [kɑ] was hardly ever misidentified (0.2 percent "t" responses), but the stop in [(ʃ)tɑ] was misidentified as "k" slightly more often (5.8 percent) than the stop in [(s)tɑ] (1.4 percent). Burstless stimuli, on the other hand, generated a large number of errors, including a small percentage (2.1) of "p" responses. The response pattern for burstless stimuli warrants some closer scrutiny; it is plotted as percent "k" responses in Figure 4, with the synthetic noise continuum along the abscissa.

The figure shows that "k" responses were more frequent to [kɑ] than to [tɑ], $F(1,9) = 120.2$, $p < .001$, and that original fricative context had an effect with [kɑ]--"k" responses being more frequent to [(ʃ)kɑ] than to [(s)kɑ]--but not with [tɑ]. This was reflected in a significant interaction, $F(1,9) = 19.7$, $p < .01$, in addition to a significant main effect of original fricative context, $F(1,9) = 41.0$, $p < .001$. However, an effect of original fricative context on [tɑ] was reflected in "p" responses (not shown in Fig. 4), which were more frequent to [(s)tɑ] than to [(ʃ)tɑ]. This pattern of results replicates Condition B.

Consider now the effect of the actual fricative noise on stop identification: The percentage of "k" responses increased significantly as the synthetic noises changed from [ʃ]-like to [s]-like, $F(8,72) = 5.8$, $p < .001$. This is the familiar effect of fricative context on stop identification. For unknown reasons, this effect was essentially restricted to [kɑ], as reflected in a significant interaction, $F(8,72) = 6.3$, $p < .001$. It is also evident that the increase for [kɑ] occurred almost exclusively on the left half of fricative noise continuum, viz., within the "sh" category. In this respect, the data replicate Experiment 4 of Mann and Repp (in press), which had combined the same synthetic fricative noises with synthetic CV portions from a [tɑ]-[kɑ] continuum. However, the present data were not sufficient to determine with any degree of confidence whether, for the ambiguous fricative noises in the middle of the [ʃ]-[s] continuum, the perceived fricative category had any separate influence on stop perception (cf. Mann & Repp, in press: Exp. 5). The present pattern of results admits that possibility; in any case, it supports our earlier conclusion (Mann & Repp, in press) that spectral properties of the fricative noise contribute significantly to the effect of the fricative on stop perception.

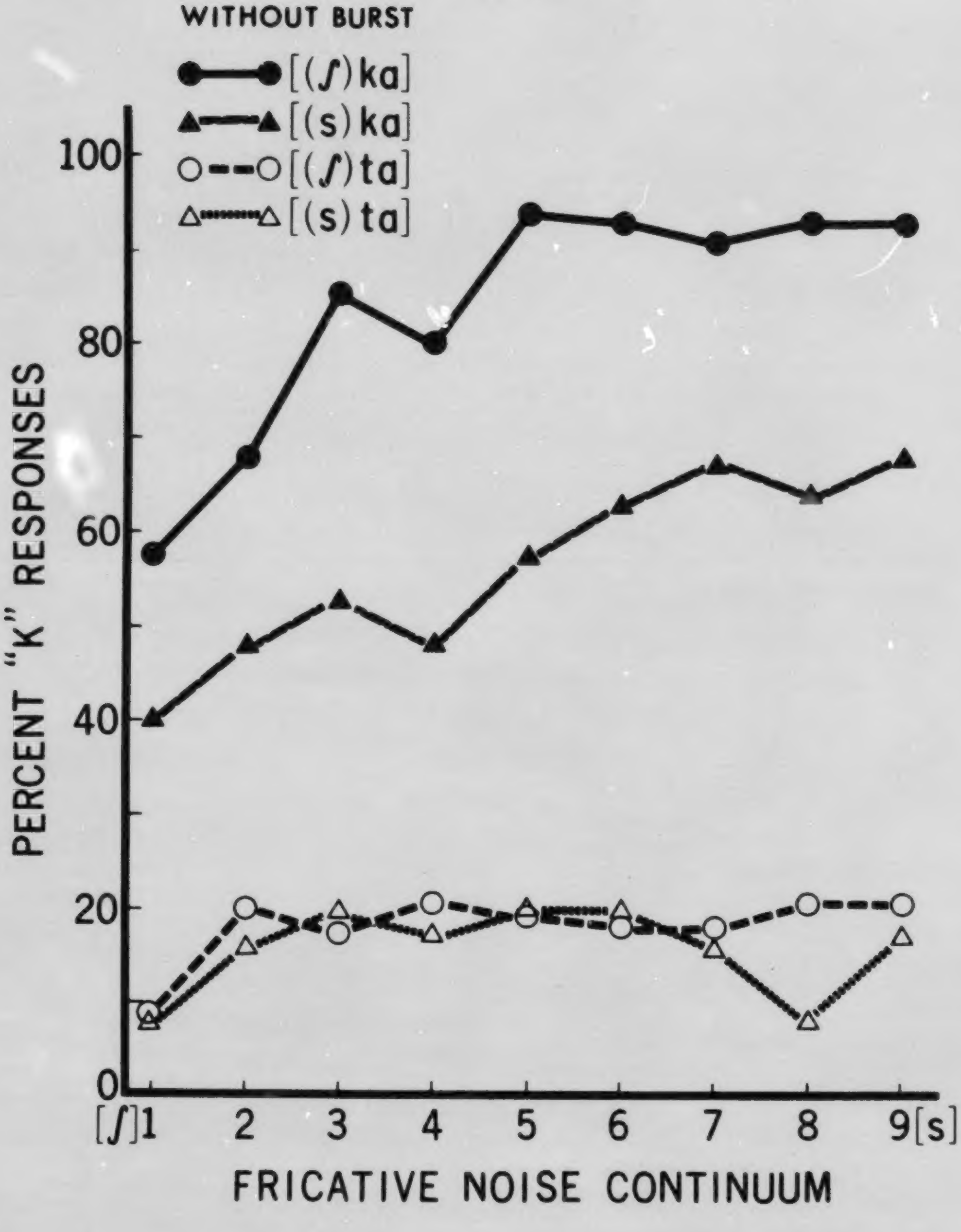


Figure 4. Effects of (synthetic) fricative noise spectrum on stop identification in CV portions excerpted from different fricative contexts (Exp. 2C).

Condition D: Fricative identification in F+(F)CV stimuli without silence

In this final experiment, we tested whether the coarticulatory traces of the original fricative context in the formant transitions (and, perhaps, in the release bursts) of the natural CV portions would influence the identification of preceding ambiguous fricative noises in a situation where the transitional cues are not interpreted as cues to place of articulation of a stop consonant (as in Condition C) but are integrated with the fricative noise cue into the fricative percept. We intended to achieve this condition by eliminating the silent interval between fricative noise and CV portion, which is a major cue for stop manner. If CV formant transitions following [s] convey a more forward place of (stop) articulation, they should, when interpreted as cues to fricative place of articulation, bias fricative perception in a more forward direction (i.e., towards "s") than transitions following [ʃ]. In the same vein, [tɑ] transitions should bias fricative perception more towards "s" than [kɑ] transitions, for [t] has a more forward place of articulation than [k].

Method. The stimulus sequence was the same as in Condition C, except that the 75-msec gap was deleted from all stimuli, so that the CV portion immediately followed upon the fricative noise. The same subjects as in Conditions A and B participated. Their task was to identify the fricative as "sh" or "s" and, *if* they heard a stop following it, to identify it as "p", "t", or "k".

Results and discussion. One very clear-cut result that had not really been expected was that all subjects heard stop consonants in the stimuli with bursts. (99.97 percent stop responses.) Thus, a silent interval was not needed to cue stop manner in this case; the presence of the release burst (plus some aspiration) was perfectly sufficient. Burstless stimuli, on the other hand, were predominantly perceived as fricative-vowel syllables, with the exception of two subjects (both paid volunteers) who reported stop consonants in these stimuli as well. For these two subjects, the percentages of stop responses to burstless stimuli were 87.5 and 99.5, respectively; the average percentage for the remaining ten subjects was 3.3. Thus, these other subjects presumably interpreted the CV formant transitions as cues to fricative place of articulation.

The fricative identification results are shown in Figure 5. The left panel shows the results for burstless stimuli. It can be seen that both predicted effects were obtained: "sh" responses were more frequent in the presence of [kɑ] transitions, $F(1,11) = 17.5$, $p < .01$, and when the transitions had originally been preceded by [ʃ], $F(1,11) = 16.8$, $p < .01$. However, both effects were primarily due to the [(ʃ)kɑ] stimuli, as confirmed by a significant interaction, $F(1,11) = 18.0$, $p < .01$.

The right panel shows the results for stimuli with bursts. Clearly, the pattern was different here: [kɑ] transitions led to *fewer* "sh" responses than [tɑ] transitions, $F(1,11) = 15.3$, $p < .01$, and there was little effect of original fricative context.[6] Thus, when the transitional cues of the CV portion were not integrated into the fricative percept but served as cues to stop identity, we obtained the retroactive context effect also found in Condition C: "sh" responses were more frequent in conjunction with "t"

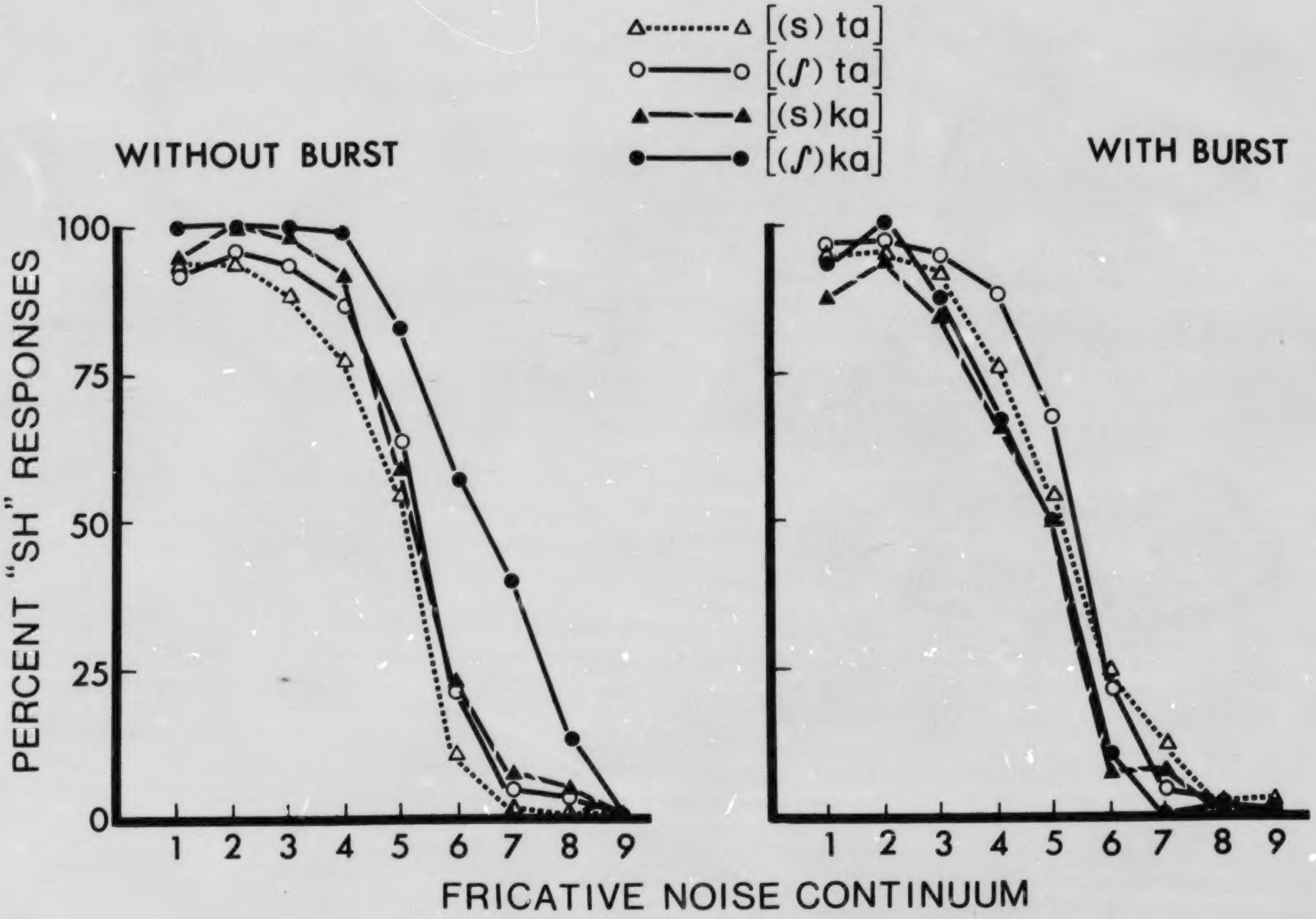


Figure 5. Effects of following CV portions on fricative perception in the absence of a stop closure interval (Exp. 2D).

responses than in conjunction with "k" responses--a contrastive retroactive effect that is complementary to the proactive effect of fricative context on stop identification.

Indeed, that familiar proactive context effect could also be observed in this experiment, viz., in the subjects' identifications of the stop consonants (if perceived). These data are summarized in Table 5. The table shows, for burstless stimuli, percentages of "t" and "k" responses contingent on whether the fricative noise was identified as "sh" or as "s". (The two subjects who gave predominantly stop responses are not included here; "p" responses did not occur.) In the two right-hand columns, the percentages of "k" responses to burstless stimuli are further conditionalized on the occurrence of a stop response, thus making them comparable to the corresponding percentages for stimuli with bursts (which always led to stop percepts).

Table 5

Stop identification, contingent on fricative identification (Exp. 2D)

<table>
<tr><th>Stimulus</th><th colspan="6">Response (percent)</th></tr>
<tr><th></th><th>"t"|"sh"</th><th>"t"|"s"</th><th>"k"|"sh"</th><th>"k"|"s"</th><th>"k"|"sh"</th><th>"k"|"s"</th></tr>
<tr><th></th><th></th><th></th><th></th><th></th><th colspan="2">(given a stop response)</th></tr>
<tr><td>Without burst</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>F+[(s)tɑ]</td><td>9.3</td><td>3.8</td><td>0.4</td><td>---</td><td>4.1</td><td>---</td></tr>
<tr><td>F+[(ʃ)tɑ]</td><td>5.9</td><td>3.4</td><td>---</td><td>0.4</td><td>---</td><td>10.5</td></tr>
<tr><td>F+[(s)kɑ]</td><td>5.8</td><td>6.4</td><td>---</td><td>4.0</td><td>---</td><td>38.5</td></tr>
<tr><td>F+[(ʃ)kɑ]</td><td>1.5</td><td>8.4</td><td>0.7</td><td>25.1</td><td>31.8</td><td>74.9</td></tr>
<tr><td>With burst</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>F+[(s)tɑ]</td><td></td><td></td><td></td><td></td><td>---</td><td>0.9</td></tr>
<tr><td>F+[(ʃ)tɑ]</td><td></td><td></td><td></td><td></td><td>2.5</td><td>5.9</td></tr>
<tr><td>F+[(s)kɑ]</td><td></td><td></td><td></td><td></td><td>99.7</td><td>100.0</td></tr>
<tr><td>F+[(ʃ)kɑ]</td><td></td><td></td><td></td><td></td><td>99.7</td><td>100.0</td></tr>
</table>

The error pattern shown in Table 5 makes good sense in the light of our previous results. The stops in stimuli with bursts were generally identified correctly, especially [kɑ]. Misidentifications of [tɑ] as "k" were more frequent when the original fricative had been [ʃ] (rather than [s]) and when the actual fricative was identified as "s" (rather than "sh"), both effects being in the expected direction. When stops were heard in burstless stimuli, it was [tɑ] that was generally identified correctly, whereas [kɑ] was actually more often labeled "t" than "k". Again, however, "k" responses were much more frequent when the original context had been [ʃ] (rather than [s]) and when the actual fricative was identified as "s" (rather than "sh").7

In summary, Condition D once more demonstrated all the previously observed effects. Listeners heard more instances of "k" when the preceding fricative was labeled as "s" (perceptual context effect). Formant transitions that had originally been preceded by [s] elicited more "s" responses (coarticulatory effect), given that the noise and transition cues were integrated, i.e., given that no stop percept intervened. Under the same conditions, [tɑ] transitions led to more "s" responses than [kɑ] transitions. If a stop was heard, the effect of original fricative context ceased, and more "s" responses were given in conjunction with "k" than with "t" (retroactive context effect). These results not only replicate the reciprocal contingency of fricative and stop identification, but also confirm once more the existence of coarticulatory traces of preceding fricatives in the formant transitions of the following signal portion.

SUMMARY AND CONCLUSIONS

The present series of experiments increases our understanding of the perceptual context effect discovered by Mann and Repp (in press)--the tendency to perceive velar stops following [s]. The effect itself must be considered firmly established, as it has been obtained consistently not only in all-synthetic stimuli (Mann & Repp, in press) but also in combinations of natural fricative noises with synthetic CV portions (Exps. 1A and 1C) and in combinations of synthetic fricative noises with natural CV portions (Exps. 2B, 2C, and 2D).

Experiment 1C successfully ruled out the hypothesis that the context effect is due to supposedly neutral fricative noises acting as direct cues to stop place of articulation. While there are demonstrable perceptual effects of direct place cues in the fricative noise, these effects are independent of the influence of fricative identity on stop perception. Our results also ruled out the possibility that a simple bias to respond "k" in conjunction with "s" underlies the context effect (Exp. 1B).

In Experiment 2, we obtained clear evidence that fricative articulation effects perceptually significant changes in the following CV portions. Thus, we have established an empirical basis for the hypothesis that the perceptual context effect represents a form of compensation for coarticulatory shifts. It is true that our data reflect the articulation of only a single speaker; it remains to be seen whether fricative-stop coarticulation is a universal phenomenon. At the very least, however, our data show that such coarticulation *can* occur.

We are aware, of course, that the demonstration of coarticulatory interactions between fricative and stop production by no means proves that they are the *cause* of the corresponding perceptual effect. Indeed, the perceptual effect may represent a general tendency to differentiate successive phonetic segments on the place-of-articulation dimension--a tendency that would parallel the general assimilatory nature of coarticulation but may not be related to the specific coarticulatory interactions between the segments in question. Experiments to prove a specific connection between perception and production are difficult to design but perhaps not impossible, and we are presently giving this issue a good deal of thought.

Our studies leave open several additional questions about the nature of the context effect of interest. For example, there is the question of whether the effect of the fricative on the stop is a function of perceived fricative category or of fricative noise spectrum. Our earlier experiments (Mann & Repp, in press) suggested that both factors are involved, and our present Experiment 2C reaffirmed a strong role of fricative noise spectrum. To the extent that future studies will replicate an effect of perceived fricative category, two separate mechanisms may be needed to explain the perceptual context effect. Perhaps, both mechanisms serve to compensate for coarticulatory effects; but it is conceivable that only one of them does.

The perceptual context effect and the associated coarticulatory shifts demonstrated here are by no means isolated or exotic phenomena. Just as coarticulation between successive phonetic segments is probably even more common than the considerable available evidence suggests, perceptual context effects appear to be the rule rather than the exception. For example, stop perception is affected not only by preceding fricatives but also by liquids (Mann, in press) and other stops (Repp, 1978). There are not only proactive context effects in perception but also retroactive ones, such as the influence of following vowels on fricative perception (Mann & Repp, in press). The parallel to the well-known bidirectionality of coarticulation is obvious. We believe that, as the evidence for perceptual and articulatory interdependencies between phonetic segments continues to increase, static and mechanistic approaches to the problem of speech perception--still in vogue but beset with increasing difficulties--will have to make way for more dynamically oriented theories.

APPENDIX

Here we report acoustic measurements of the natural-speech stimuli used in our experiments. All spectral measurements were made by visual inspection of successive spectral cross-sections, provided by a Federal Scientific UA-6A spectrum analyzer and displayed as point plots on a Hewlett-Packard 1300A scope. All spectra were computed over 25.6-msec windows in 12.8-msec steps; they were smoothed and pre-emphasized. Maximum resolution was 40 Hz. The precise position of the windows with respect to stimulus onset (or offset) could not be controlled; we simply took the first (last) cross-section that yielded clear spectral peaks as the stimulus onset (offset). The effect of this uncertainty in temporal alignment on the measurements was considered negligible.

Fricative Noises

There were 18 stimuli to be measured: three tokens of [ʃ] noise and three tokens of [s] noise from each of three original contexts, [-ɑ], [-tɑ], and [-kɑ]. We examined the last 10 sections (128 msec) of each stimulus, starting with the last and proceeding backwards. From each spectrum, we determined the location of the major energy peaks below 5 kHz, as well as the lower cutoff frequency--the point below which there was either no energy at all or only small, isolated peaks. (This latter measure may have been dependent on input amplitude and therefore should not be taken absolutely; however, it is highly relevant to a comparison of noises from different contexts.) Having determined these measures, we averaged them across the three tokens of each noise in each context, omitting all values that were spurious or inconsistent within or across tokens. A graphic representation of these average parameters for [ʃ] and [s] noises in [-(tɑ)] and [-(kɑ)] context is provided in Figure 6.

Figure 6 represents spectral peaks as connected circles and the lower cutoff frequencies as simple lines. The figure shows that the major resonances (poles) were fairly steady-state and not much influenced by the place of stop occlusion. Obviously, the parameter sensitive to stop occlusion was the lower cutoff frequency, particularly in the last 50 msec. In the context of [tɑ], the lower edge of the spectrum shifted rapidly upward, whereas, in the context of [kɑ], [s] showed a small downward shift, and [ʃ] showed a large downward shift followed by a small upward shift. At stimulus offset, the cutoff frequencies differed by 600-800 Hz between [-(tɑ)] and [-(kɑ)] stimuli. In addition, tokens of [s(kɑ)] showed scattered patches of energy below the cutoff frequency over the last 50 msec; if those peaks, one of which was as low as 300 Hz (not shown in Fig. 6), had been included in the cutoff frequency estimate, the dip in the cutoff function for [s(kɑ)] in Figure 6 would, of course, have been much more dramatic. There is an indication in Figure 6 that the earlier portion of the [ʃ] noise was also affected by context: In [ʃ(kɑ)], but not in [ʃ(tɑ)], there was initially an energy minimum between the two lower spectral peaks.

Tokens of [s(ɑ)] and [ʃ(ɑ)]--not shown in Figure 6 for reasons of clarity--were highly similar in spectral structure to the other noises, except that they did not show any pronounced changes in lower cutoff frequency at offset. Their average cutoffs at offset were just about halfway between those for [-(tɑ)] and [-(kɑ)] stimuli.

Thus, our data suggest that fricative noises preceding a stop closure are characterized by a rapid loss of low-frequency energy preceding [t] and by a relative increase in low-frequency energy preceding [k], these changes taking place within the last 50 msec or so. The major spectral peaks, on the other hand, do not seem to shift with place of stop occlusion, at least in the range below 5 kHz. Since our observations are based on a very small number of utterances of a single speaker, we should not draw any conclusions except that we have described the acoustic basis for the perceptual effects observed in Experiments 1B and 1C. However, our data seem to agree with earlier informal reports in the literature (Malécot & Chermak, 1966; Uldall, 1964).

The durations of our fricative noises (averaged across tokens) were as follows: [s(ɑ)], 211 msec; [ʃ(ɑ)], 216 msec; [s(tɑ)], 208 msec; [s(kɑ)], 204 msec; [ʃ(tɑ)], 158 msec; [ʃ(kɑ)], 157 msec. Thus, it appears that our speaker shortened his [ʃ] noises considerably more than his [s] noises when a stop consonant followed.

CV Portions

For each of the 12 CV portions (3 tokens each of [(s)tɑ], [(ʃ)tɑ], [(s)kɑ], and [(ʃ)kɑ]), we traced the major spectral peaks (formants) through the first 10 spectral sections that yielded a clear formant pattern. Thus, we did not include the release burst whose spectrum was too irregular (especially

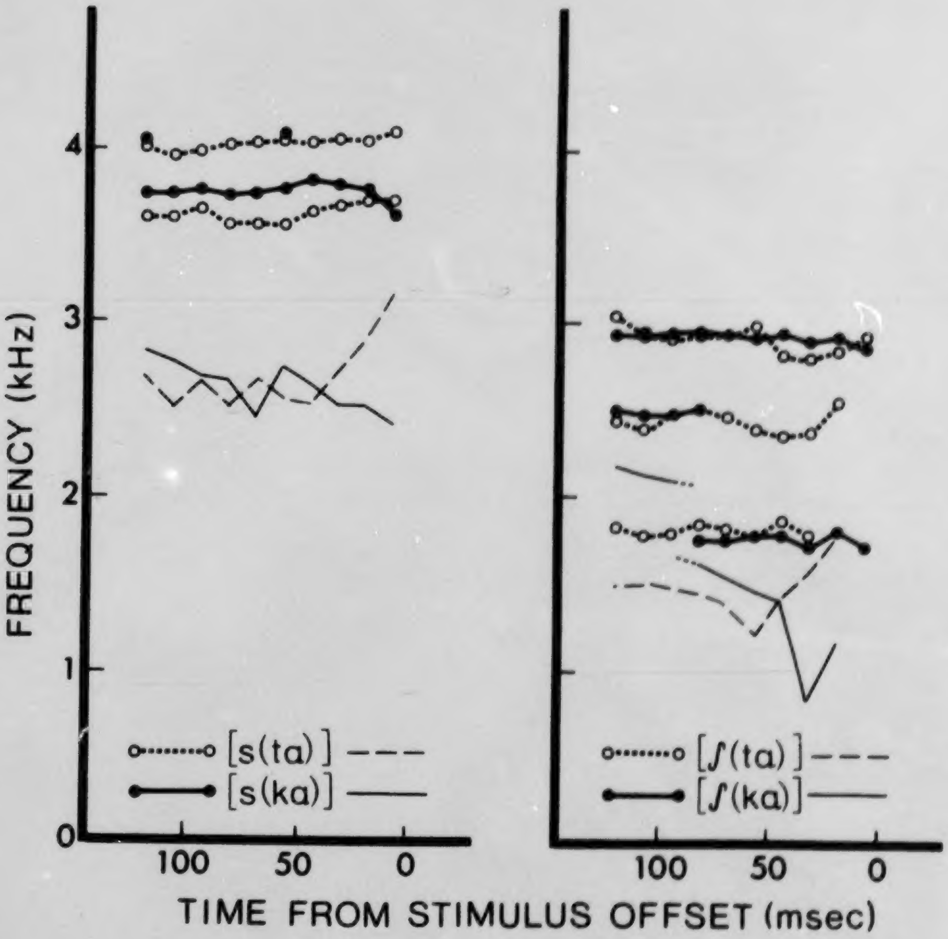


Figure 6. Average spectral structure of fricative noises in different stop contexts (Exp. 1).

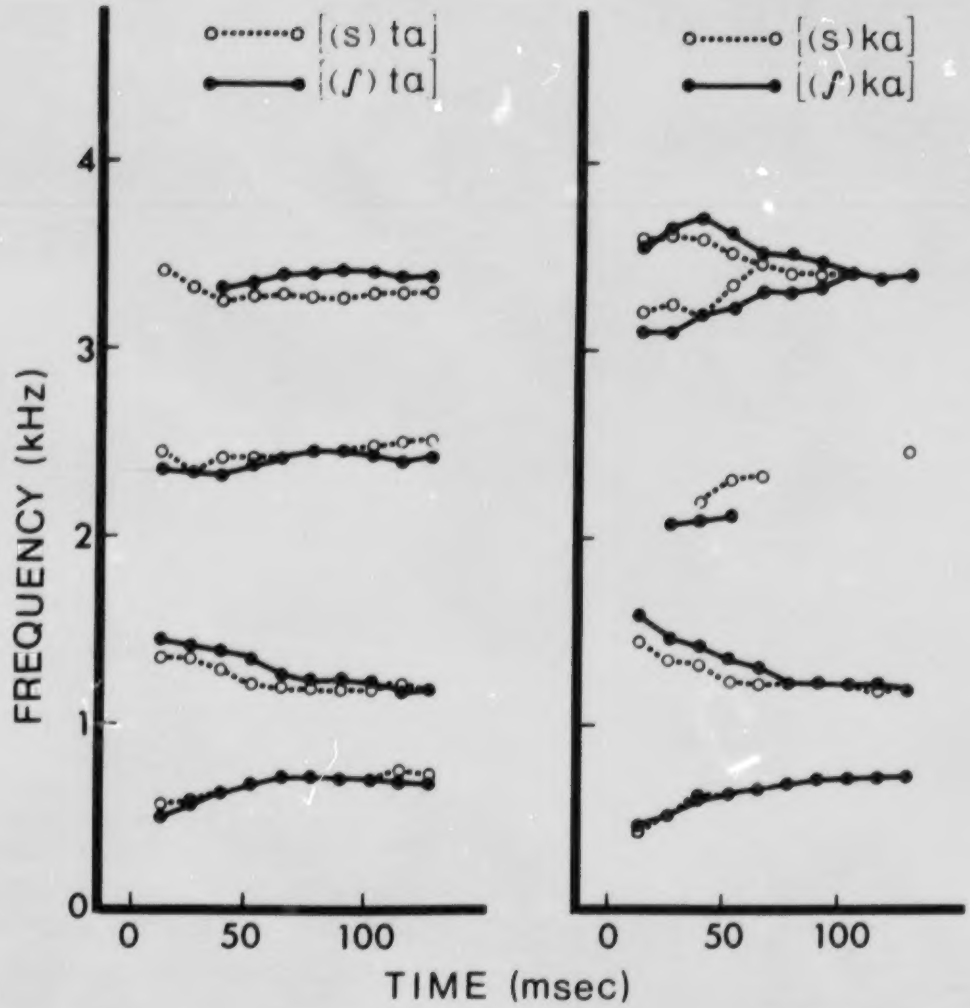


Figure 7. Average spectral structure of CV portions in different fricative contexts (Exp. 2).

in [tɑ]) to permit useful comparisons, given the limited amount of data. The formant trajectories, averaged across tokens, are displayed in Figure 7.

It can be seen that, although there had been clear perceptual differences between (burstless) CV stimuli from different fricative contexts, the acoustic effects of the preceding fricative were rather small: The second formant had a somewhat higher frequency (by up to 100 Hz) following [ʃ] than following [s], and this difference seemed to persist throughout the transitional phase (about 50 msec). There are indications of a higher onset of F3 in [(s)kɑ] than in [(ʃ)kɑ], but this formant was weak and often altogether absent in [k]. The differences observed, though small, are in agreement with a forward shift in place of stop occlusion following [s], since a forward shift implies a greater separation of F2 and F3 onsets (cf. the greater separation for [tɑ] than for [kɑ]). The "split F4" for [kɑ] appears to be an idiosyncratic feature of the speaker who produced these utterances.

We have examined a larger corpus of utterances from several speakers and, so far, have not found consistent evidence for coarticulatory shifts in CV formant transitions following [s] vs. [ʃ]. If these shifts exist--as our experimental utterances suggest--they must be rather small. It is also possible, of course, that not all speakers coarticulate stops with preceding fricatives. We are continuing our investigations in that direction.

The durations of our CV stimuli ranged from 440 to 540 msec, although the major energy was contained within the first 300 msec or so. The durations of the burst-cum-aspiration portions--which were removed to obtain the burstless versions--varied from 18 to 33 msec. On the average, they were slightly longer for [kɑ] (25.2 msec) than for [tɑ] (21.5 msec); there was little difference between fricative contexts.

We did not measure stimulus amplitudes since an earlier study of ours (Mann & Repp, in press: Exp. 4) suggested that the relative amplitude levels of the noise and CV portions have little influence on perception. Suffice it to say that, when substituting synthetic for natural stimulus portions, we tried to maintain approximately the original amplitude relationships.

REFERENCES

Ali, L., Gallagher, T., Goldstein, J., & Daniloff, R. Perception of coarticulated nasality. *Journal of the Acoustical Society of America*, 1971, *49*, 538-540.

Bailey, P. J., & Summerfield, Q. Information in speech: Observations on the perception of [s]+stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 536-563.

Benguerel, A. P., & Adelman, S. Coarticulation of lip rounding and its perception. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception*. New York: Springer-Verlag, 1975.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 1979, *65*, 1518-1532.

Fant, G., Liljencrants, J., Malac, V., & Borovickova, B. Perceptual evaluation of coarticulation effects. *Quarterly Progress Report on Speech Research* (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm), 1970, *1*, 10-13.

Lehiste, I., & Shockey, L. On the perception of coarticulation effects in English VCV syllables. *Working Papers in Linguistics* (Department of Linguistics, Ohio State University), 1972, *12*, 78-86.

Malécot, A., & Chermak, A. Place cues for /p.t.k./ in lower cut-off frequency frequency shifts of contiguous /s/. *Language and Speech*, 1966, *9*, 162-169.

Mann, V. A. Influence of preceding liquid on stop perception. *Perception & Psychophysics*, in press.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 1980, *28*, 213-228.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, in press.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, 1978, *24*, 471-485.

Schwartz, M. F. Transitions in American English /s/ as cues to the identity of adjacent stop consonants. *Journal of the Acoustical Society of America*, 1967, *42*, 897-899.

Uldall, E. Transitions in fricative noise. *Language and Speech*, 1964, *7*, 13-14.

Yeni-Komshian, G., & Soli, S. Extraction of vowel information from fricative spectra. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th Meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

FOOTNOTES

[1] Perceptual coherence between synthetic and natural signal portions is not difficult to achieve, especially when--as in the present studies--they have different sources of excitation (aperiodic fricative noise vs. largely periodic CV portion) and, moreover, are separated by a silent closure interval. However, we have also been successful in combining natural and synthetic voiced portions, separated by silence (Mann, in press), and synthetic noises with natural voiced portions, immediately adjoined (Mann & Repp, 1980).

[2] Errors in fricative identification were virtually nonexistent, except for a single subject (14 percent) whose exclusion would not have changed the results.

[3] Only two subjects made any errors in fricative identification (2 and 7 percent, respectively).

[4] As in Condition A, only one subject committed a large number of fricative identification errors (22 percent); nevertheless, he showed the pattern of Figure 2. Exclusion of his data would not have changed the results.

[5] The result was unexpected because velar stops were thought to be more susceptible to articulatory shifts than alveolar stops. However, this hypothesis is difficult to test perceptually, because the probability of confusions along the place dimension depends on the perceptual distances between the few alternative categories available. Most likely, "th" is closer to "d" than "d" is to "g". Therefore, a small forward shift in the articulation of [tɑ] will result in a large number of "th" responses, whereas a larger forward shift in the articulation of [kɑ] might result in only a moderate number of "d" responses. As will be seen, omission of the "th" category in Condition B led to the "expected" better identifiability of alveolar stops.

[6] Inspection of the data of the two subjects who had heard stops in burstless stimuli--and who are included in Figure 5--revealed that they showed only minimal effects of burstless CV portions on fricative identification. It was interesting to note that one of these subjects identified all stops as "t" while the other alternated fairly randomly between "t" and "k", both being atypical response patterns suggesting that these listeners did not process the transitional cues properly.

[7] The combination of these two factors also elicited the largest absolute number of stop responses (25 percent), probably due to the incompatibility of [(ʃ)kɑ] transitions with an [s]-like fricative noise. These stop responses derived almost exclusively from the two most [s]-like noises (stimuli 8 and 9 on the fricative noise continuum). A curious and not fully explained finding was a greatly increased percentage of stop responses (20 percent) following the most ambiguous fricative noise (No. 5 on the [ʃ]-[s] continuum). Perhaps, the relative inappropriateness of that noise for either fricative category obviated its perceptual integration with the following vocalic formant transitions.

INFLUENCE OF VOCALIC CONTEXT ON PERCEPTION OF THE [ʃ]-[s] DISTINCTION: IV. TWO STRATEGIES IN FRICATIVE DISCRIMINATION

Bruno H. Repp

Abstract. Synthetic noises from a [ʃ]-[s] continuum, followed by vocalic portions known to influence the location of the [ʃ]-[s] boundary, were presented in AXB and fixed-standard AX discrimination tasks. The majority of naive subjects perceived these fricative-vowel syllables fairly categorically in both tasks; that is, discrimination performance followed the patterns predicted from identification scores, including shifts contingent on the nature of the vocalic portion. However, two subjects achieved much better discrimination scores than the rest; their results were similar to those of three experienced listeners who participated as additional subjects in the AX task. Most significantly, influences of vocalic context for these listeners were either absent or reversed in direction relative to the effects shown by the categorical perceivers. However, all listeners showed regular context effects in a phonetic labeling task. These results are consistent with the view that influences of vocalic context on fricative identification are tied to phonetic perception--they disappear in listeners who (judging from their much better performance) are successful in following the nonphonetic strategy of restricting attention to the spectral properties of the fricative noise portion.

INTRODUCTION

Several recent studies (Mann & Repp, 1980; Whalen, in press; Kunisaki & Fujisaki, Note 1) have shown that perception of the [ʃ]-[s] distinction is sensitive to the nature of the subsequent vocalic context. Two separate effects may be distinguished. One is due to the quality of the following vowel: Given a somewhat ambiguous fricative noise (often a necessary condition for observing any contextual effects--cf. Harris, 1958), listeners tend to perceive "s" in the context of a rounded vowel (such as [u]) but "sh" in the context of an unrounded vowel (such as [ɑ]). The other effect is due to the nature of the vocalic formant transitions: Listeners tend to perceive "s" when the transitions resemble those normally following [s] frication, and "sh" when the transitions resemble those normally following [ʃ] frication. The

---

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories. I would like to thank Patti Price for help in conducting the experiments, and Virginia Mann and Doug Whalen for fruitful discussions and for serving as "expert listeners" in Experiment 2. Other colleagues who have contributed valuable comments include Catherine Best, Alvin Liberman, and David Pisoni.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

vowel quality and transition effects are both reliable and pronounced, especially when synthetic fricative noises are spliced together with natural-speech vocalic portions (Mann & Repp, 1980; Whalen, in press).

One important theoretical question raised by these findings is whether the effects of vocalic context on fricative perception arise at a phonetic (speech-specific) level of processing, or whether they are due to some auditory interaction between adjacent stimulus segments. Even though what is known about other contextual effects in speech perception generally suggests a phonetic origin, evidence supporting this contention needs to be adduced for each individual effect, considering the large number of possible auditory interactions and the sizeable group of researchers who seem to believe that such interactions can explain most or all phenomena in speech perception.

Consider first the transition effect. If it is phonetic in nature, it is best described as resulting from the perceptual integration of two separate cues--the fricative noise and the following formant transitions--into a single phonetic percept. The integration is motivated by the fact that both noise and transitions are necessary consequences of producing either [s] or [ʃ]. On the other hand, if the effect is auditory in origin, it seems implausible that it would arise from perceptual integration, considering the great spectral disparity of the two cues. Rather, the assumptions would be that listeners focus on one cue only (most likely on the noise portion) and that the perception of the relevant auditory properties of the fricative noise is somehow modified by the formant transitions (or vice versa). The auditory mechanisms that could mediate such a perceptual interaction are not obvious, but auditory contrast and nonsimultaneous masking are candidates.

Consider now the vowel quality effect. A phonetic explanation for listeners' tendency to hear "s" rather than "sh" in the context of rounded vowels appeals to a well-known coarticulatory effect: Fricative noises preceding rounded vowels characteristically exhibit a downward shift in spectrum, due to anticipatory lip rounding (Mann & Repp, 1980; Kunisaki & Fujisaki, Note 1). Thus, listeners appear to compensate in perception for a consequence of coarticulation. Of course, such compensation could never occur at a level of processing that has no access to tacit knowledge of articulatory dynamics and contextual variations in speech cues. Therefore, to explain the vowel quality effect in auditory terms, we must again assume that the auditory percept of the fricative noise is somehow influenced by the following signal portion (e.g., through some form of spectral contrast).

Since the formant transitions are acoustically dependent on vowel quality, the auditory hypothesis thus attempts to explain both vowel quality and transition effects by essentially the same mechanism--an auditory effect of the vocalic onset spectrum on perception of the fricative noise. Thus, this hypothesis has the advantage of parsimony; as we have seen, the vowel quality and transition effects have quite different explanations in a theory of phonetic perception--explanations that are united only by their common appeal to articulatory dynamics as a perceptual guideline.

The present study was conducted to answer the following question: If listeners are led by the task demands to focus on the spectral quality of the fricative noise rather than on its phonetic category, would their responses still be influenced by the periodic stimulus portion following the noise? Presumably, a strictly auditory effect of vocalic context on fricative noise perception would operate whether or not listeners restrict their attention to the noise portion alone. In fact, such a focusing of attention is already implied in the auditory hypothesis, and a further effort on the listener's part should have little if any effect. On the other hand, if the effects of vocalic context are phonetic in nature, they might disappear when listeners focus on the auditory quality of the noise portion, i.e., when they use a perceptual strategy that presumably bypasses the mechanisms specific to phonetic perception.

The extent to which listeners would be successful in adopting such a nonphonetic strategy in judging fricative-vowel stimuli was not known in advance. Many speech stimuli are categorically perceived; that is, untrained listeners perceive the stimuli in terms of phonetic categories even when attempting to make fine auditory discriminations. Typically, however, stimuli that are categorically perceived are distinguished by rather subtle acoustic differences that can be detected only by trained listeners (see, e.g., Carney, Widin, & Viemeister, 1977; Edman, 1979). Fricative-vowel syllables, on the other hand, contain a prolonged noise portion, and it would seem that listeners should be able to detect (sufficiently large) differences in the noise without too much difficulty.[1] Certainly, isolated noises from a [ʃ]-[s] continuum can be discriminated quite easily, even though they can also be labeled phonetically as "sh" or "s" (Healy & Repp, 1980).

In the present study, two different discrimination paradigms were used (AXB and fixed-standard AX) which were expected to differ in the extent to which they facilitated the task of making fine auditory discriminations in the noise portion (cf. Creelman & Macmillan, 1979). In both discrimination tests, fricative noises from a [ʃ]-[s] continuum were followed by several different vocalic portions. An initial identification test was expected to confirm the earlier finding (Mann & Repp, 1980) that the [ʃ]-[s] boundary shifts with a change in vowel quality or formant transitions. The central question was whether analogous shifts would be observed in the discrimination tasks (as predicted if the stimuli are categorically perceived) or whether selective attention to the auditory properties of the noise portion, especially in the sensitive fixed-standard AX test, would result in a disappearance of vowel context effects.

EXPERIMENT 1: IDENTIFICATION AND AXB DISCRIMINATION

Method

Subjects. Eight paid student volunteers participated. None of them was experienced in speech discrimination tasks, although some of them had taken part in earlier experiments requiring identification of stimuli similar to those used here.

Stimuli. The stimuli consisted of synthetic noise portions followed by natural-speech periodic portions. The fricative noises were generated on the OVE IIIc serial resonance synthesizer at Haskins Laboratories and constituted a 9-member [ʃ]-[s] continuum. The endpoint stimuli were chosen to match approximately in spectrum (below 5 kHz) the [ʃ] and [s] noises of the speaker from whose utterances the periodic portions were taken. The frequencies of the two poles (formants) that characterized each noise are listed in Table 1. Noise duration was 200 msec; the amplitude contour peaked after 150 msec. Overall amplitude was nearly constant across the continuum.

Table 1

Fricative noise stimuli of Experiment 1
(pole center frequencies in Hz)

Stimulus No.		Pole 1	Pole 2
[ʃ]	1	2466	3108
	2	2613	3293
	3	2769	3488
	4	2933	3695
	5	3108	3915
	6	3293	4148
	7	3489	4394
	8	3697	4655
[s]	9	3917	4932

The periodic stimulus portions were excerpted from utterances of [sɑ], [ʃɑ], [su], and [ʃu], produced by a male speaker of American English. To indicate the absence of the original fricative noise (but the presence of appropriate formant transitions), these portions will be referred to as [(s)ɑ], etc. In an earlier study (Mann & Repp, 1980: Exp. 4), the very same portions had dramatic effects on fricative identification when preceded by synthetic fricative noises from a [ʃ]-[s] continuum similar to the present one. That earlier experiment used three different tokens of each periodic portion, but since token variation was small, a single token of each was deemed sufficient for the present study. Fricative-vowel syllables were constructed by immediately following a synthetic noise with a periodic portion, both having been digitized at 10 kHz and low-pass filtered at 4.9 kHz.

There were four identification tests and four AXB discrimination tests, one of each for each periodic portion (a blocked factor). Each identification test contained 10 repetitions of the 9 stimuli resulting from the 9 different noises followed by one particular periodic portion. They were arranged in 5 lists of 18, with ISIs of 3 sec. Each AXB discrimination test contained 6 repetitions of the 7 2-step comparisons (1-3, 2-4, etc.) in each of 4 AXB arrangements (AAB, ABB, BAA, BBA), resulting in 168 stimulus triads. These were arranged in 6 lists of 28, with ISIs of 500 msec within triads, 3 sec between triads, and 10 sec between lists. The first list of 28 served as practice and was not scored.

Procedure. Each AXB test was preceded by the corresponding identification test. The four conditions deriving from the four different periodic portions were distributed over two sessions in counterbalanced order. The subjects listened over TDH-39 earphones in a quiet room. The tapes were played back on an Ampex AG-500 tape recorder. In the identification task, the subjects identified the fricative in each stimulus by writing down "sh" or "s". In the AXB discrimination task, the responses were "A" or "B", depending on whether the second stimulus in a triad was judged to be the same as the first or as the third. The subjects were told to listen carefully for any difference in the noise portion, and to guess if necessary.

Results and Discussion

Identification. Although the identification test was essentially a partial replication of Experiment 4 of Mann and Repp (1980), there were two important differences: (1) The [ʃ]-[s] continuum was more realistic, the endpoints having been modeled on natural speech. (2) The different periodic portions were blocked rather than randomized. Both changes might be expected to reduce the magnitude of contextual influences on fricative perception: The improved noises were perhaps less ambiguous; and blocked presentation gave listeners an opportunity to adapt to a given periodic portion and to adjust their criteria accordingly. Therefore, it seemed important to demonstrate that vocalic context still influences fricative perception under these conditions.

The results are shown in Figure 1. It is evident that the labeling functions shifted with vocalic context in the expected directions. Listeners were more likely to perceive "sh" in the context of [ɑ] than in the context of [u], $F(1,7) = 17.1$, $p < .01$, and they were more likely to perceive "sh" when [ʃ]-transitions were present than when [s]-transitions were present, $F(1,7) = 21.2$, $p < .01$. The interaction between the vowel quality and transition effects was not significant, $F(1,7) = 0.5$, suggesting that the two effects are independent. The boundary shifts were considerably smaller in magnitude than those observed by Mann and Repp (1980), probably for both of the reasons mentioned (viz., improved fricative noises and blocked periodic portions). However, they were reliable and sufficiently large to predict shifts in discrimination peaks, if categorical perception obtains.

AXB Discrimination. Preliminary inspection of the results revealed that two of the eight subjects outperformed the others by a wide margin: Their average score was 96 percent correct. Since these two subjects apparently did something different from the rest, and since their data did not contain any information because of the ceiling effect, their results were excluded.[2] The following results are based on the remaining six subjects only.

The average discrimination functions are shown in Figure 2 separately for each periodic portion, together with predictions derived from the identification results (separately for each subject and then averaged), using the classic low-threshold model of categorical perception (Liberman, Harris, Hoffman, & Griffith, 1957; Pollack & Pisoni, 1971). It is evident that discrimination performance followed the predicted pattern quite closely, except in the [(s)u] condition where the match was less good. Discrimination was much better in the boundary region than within phonetic categories,

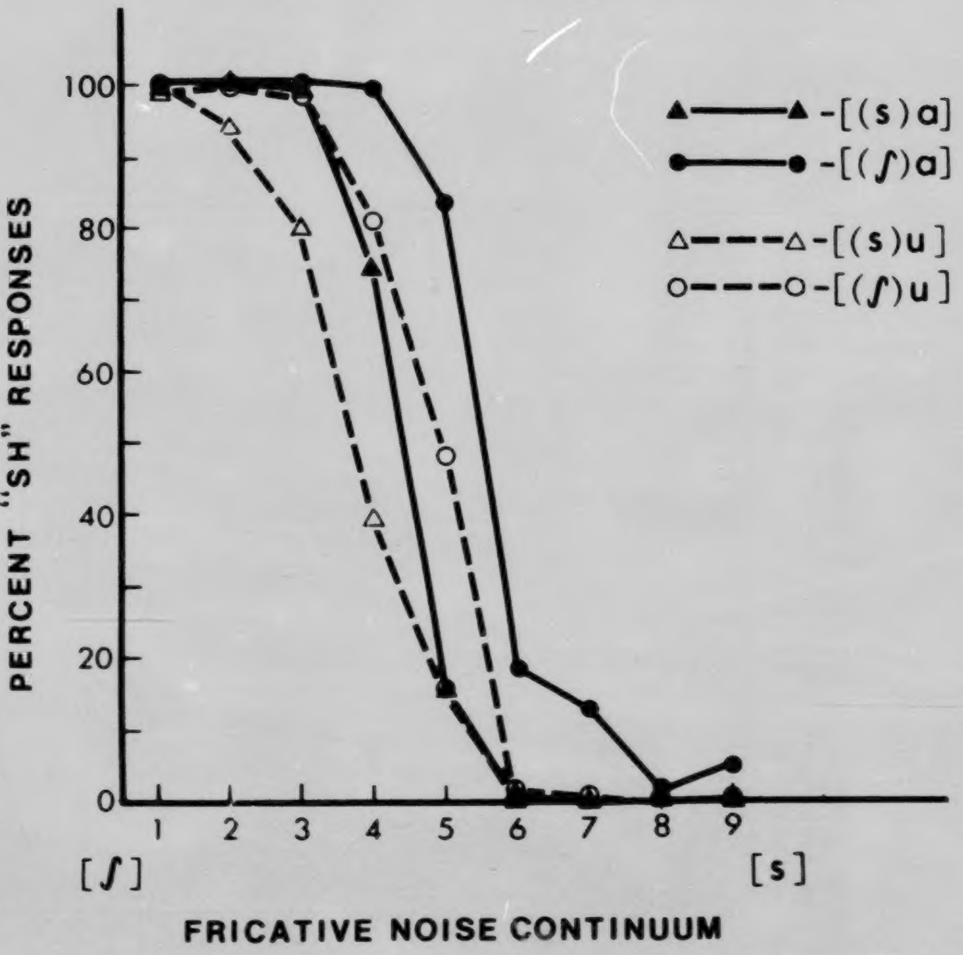


Figure 1. Identification functions for a [ʃ]-[s] noise continuum in four different vocalic contexts.

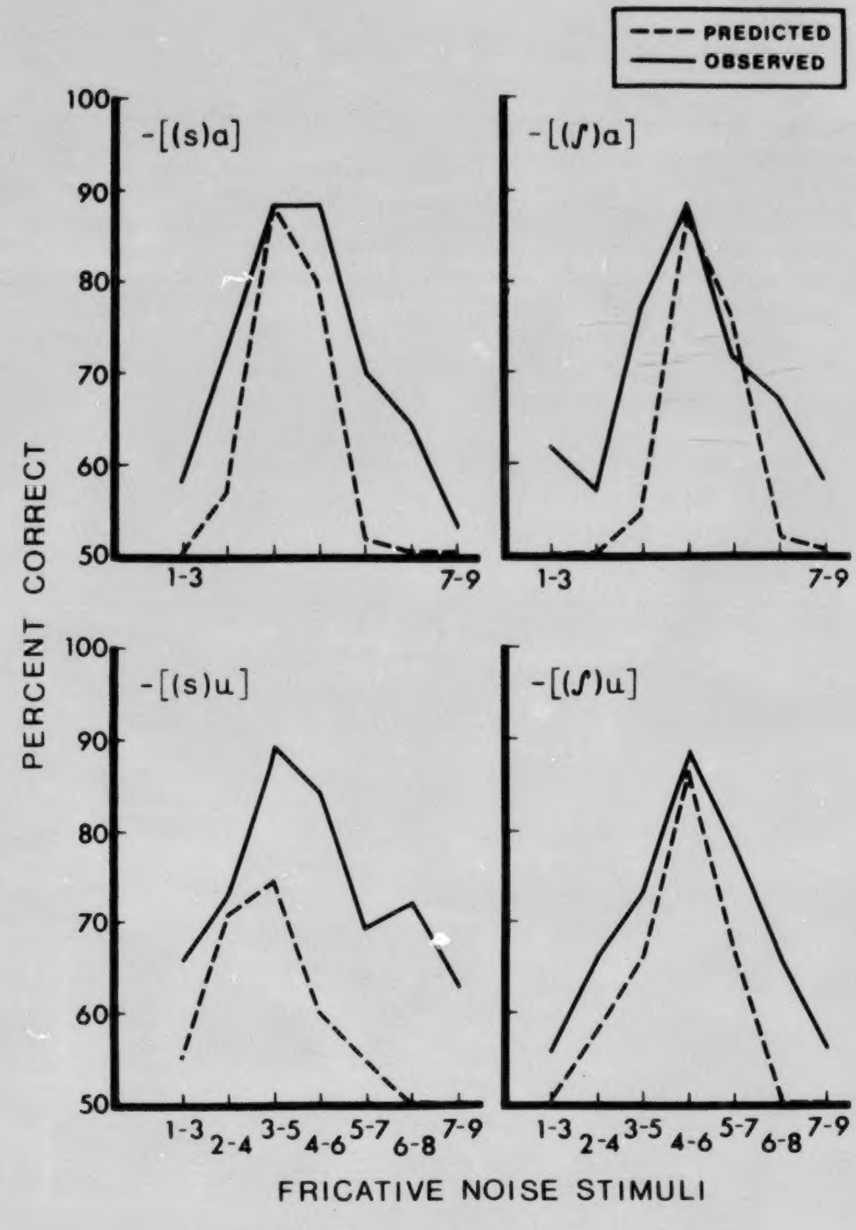


Figure 2. AXB discrimination performance of six "categorical" subjects: Two-step comparisons of [ʃ]-[s] noises in four different vocalic contexts.

although it was everywhere above chance and usually a good deal better than predicted. There were also indications that the peaks of the discrimination functions shifted as predicted with the nature of the periodic portion, although these shifts did not reach significance here because of the small number of subjects.

At least part of the difference between obtained and predicted discrimination performance may be ascribed to contrast effects in (covert) labeling during the discrimination task (Repp, Healy, & Crowder, 1979; Healy & Repp, 1980). Therefore, the results of these six subjects indicate quite strong categorical perception, in agreement with earlier findings of Fujisaki and Kawashima (Note 2) and of May (1979). Apparently, these listeners found it difficult to abandon a phonetic mode of listening and to focus on the auditory quality of the fricative noise; they seemed to make their decisions largely on the basis of the category labels, "sh" and "s". It was thought, however, that the more stringent fixed-standard AX discrimination task might lead subjects to adopt a different strategy, of the kind already evidenced by the two exceptional listeners (and by the author as a pilot subject) in the AXB task. There is little doubt that the high accuracy achieved by these latter subjects reflected a noncategorical, auditory mode of listening.

EXPERIMENT 2: FIXED-STANDARD AX DISCRIMINATION

Method

Subjects. Ten paid student volunteers participated, seven of whom had previously been subjects in Experiment 1, including the two exceptional listeners. (In addition, a panel of experienced listeners took the test--see below.)

Stimuli. Since the fixed-standard AX task was expected to facilitate discrimination, and since it had to be sufficiently difficult for even the best subjects to produce some errors, a more closely spaced 7-member fricative noise continuum was synthesized. The pole frequencies of the noises are listed in Table 2. The relationship between the two poles was somewhat different in these stimuli than in those of Experiment 1; the present stimuli were more closely related to the continuum used earlier by Mann and Repp (1980), spanning the region of highest ambiguity between [ʃ] and [s]. Only two periodic portions were used, [(ʃ)ɑ] and [(s)u], taken from Experiment 1. Thus, the vowel quality and transition effects were deliberately confounded in this study by choosing the two periodic portions that gave a maximal difference in Experiment 1.

Stimulus 4 on the noise continuum was chosen as the fixed standard. In each stimulus pair, the standard occurred first, followed by a comparison stimulus which could be any of the seven stimuli, with equal probability. Thus, only one seventh of the stimulus pairs had in fact identical noises. There were four different conditions. In two conditions, the standard and the comparison always had the same periodic portion--[(ʃ)ɑ] in one condition and [(s)u] in the other. In the other two conditions, the periodic portions were always different--[(ʃ)ɑ] for the standard and [(s)u] for the comparison in one condition, and the reverse assignment in the other.

Each condition contained 24 repetitions of the 7 possible stimulus pairs, arranged in 6 lists of 28, with ISIs of 500 msec within pairs, 2 sec between pairs, and 10 sec between lists. The first list of 28 served as practice and was not scored; thus, the results are based on 20 responses per pair per subject.

Table 2

Fricative noise stimuli of Experiment 2
(pole center frequencies in Hz)

Stimulus No.	Pole 1	Pole 2
1	2690	4030
2	2769	4148
3	2850	4269
4	2933	4394
5	3019	4523
6	3108	4655
7	3199	4792

Procedure. All four conditions were presented in a single session in counterbalanced order, with the restriction that the condition with equal periodic portions always immediately preceded the condition with the same standard but a different periodic portion in the comparison stimuli. The task was to write down "d" whenever a difference between the noises could be detected, and "s" otherwise. Guessing was discouraged. The subjects were not informed about the true frequency of identical pairs.

Results and Discussion

Even if the subjects were only moderately successful in this task, their "different" responses should show a pronounced minimum for stimulus pairs containing identical noises, and a rapid increase as a function of the physical distance of the comparison stimulus from the standard, in both directions. In other words, if listeners operate in an auditory mode, "different" responses plotted as a function of comparison stimulus number should exhibit a V-shaped pattern. Preliminary inspection of the results revealed that, surprisingly, only two out of ten listeners showed this pattern. These two subjects, whose performance was also much better, were precisely the two subjects who had performed at the ceiling in the AXB discrimination task (Exp. 1). Therefore, their data were again separated from the rest; they will be considered below.

Let us examine first the combined results of the other eight subjects, which are plotted in the top panels of Figure 3. The two conditions with identical periodic portions are on the left, those with different periodic portions are on the right. It can be seen that performance was extremely poor (a horizontal function represents chance performance), decidedly asymmetric around the standard (stimulus No. 4), and strongly influenced by the nature of the periodic stimulus portion. Comparison of the two figure panels suggests that it was the periodic portion of the standard stimulus, rather than that of the comparison, that determined the shape of the response function; this effect (the standard-periodic-portion by stimulus number interaction) was highly significant, $F(6,42) = 7.5$, $p < .001$. There tended to be more "different" responses when the periodic portions in a pair were different than when they were the same, $F(1,7) = 5.0$, $p < .10$.

How is this pattern of responses to be interpreted? Clearly, it is not random, despite the poor performance. The most obvious possibility is that these subjects remained in a phonetic mode, despite instructions to focus on the noise and despite a fixed-standard paradigm, which should have facilitated the task. What would the categorical predictions look like in this paradigm? A difficulty arises here, because no identification data were collected for the stimuli used in this experiment. Although similar stimuli had been used by Mann and Repp (1980: Exp. 4), calculations showed that the effects of vocalic context in that study were much too large to generate good predictions of the present data. The smaller stimulus range used here, together with the particular format of presentation, may of course have modified the magnitude of context effects. Therefore, hypothetical identification functions were generated on paper by trial and error to see whether predictions could be derived that resembled the results in Figure 3. This exercise had some success: If a sufficiently small effect of vocalic context is assumed (a separation of -[(ʃ)a] and -[(s)u] identification functions by about two steps on this closely spaced fricative continuum), the resulting predictions of AX performance do show the characteristic crossed pattern of the functions in the top panels of Figure 3; they also exhibit the increased rate of "different" responses in the right panel as compared to the left. However, there were also some discrepancies. Of course, the procedure of estimating labeling functions from discrimination data (rather than the other way around) is fraught with problems: It does not consider the likely occurrence of contrast effects in the AX paradigm (cf. Repp et al., 1979; Healy & Repp, 1980) and the equally likely availability to listeners of some amount of auditory information beyond the phonetic categories. However, the predicted pattern was sufficiently similar to the obtained pattern to lend plausibility to the claim that this group of subjects remained essentially in a categorical (phonetic) mode of perception even in the fixed-standard AX task. Certainly, the pattern of results cannot be explained simply as resulting from poor auditory discrimination performance; in that case, the discrimination functions should have been more clearly V-shaped.

Consider now the results of the other subjects. As mentioned above, two subjects performed much better than the rest. Their data were augmented by those of three experienced listeners--the author and two colleagues, both of whom are involved in related research on fricative perception. The average results of all five subjects are shown in the bottom panels of Figure 3. Here we see the expected V-shaped pattern: "Different" responses were least

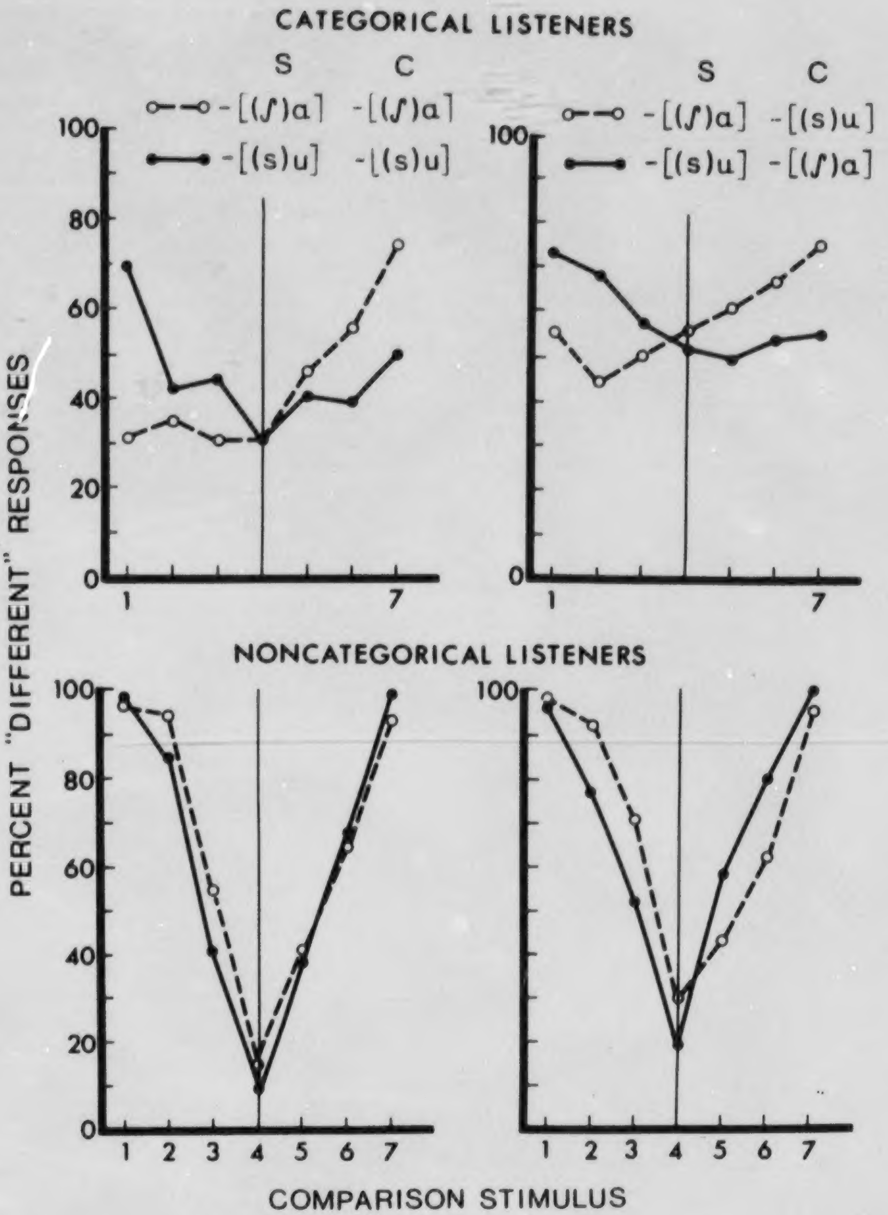


Figure 3. Fixed-standard AX discrimination performance of eight "categorical" subjects (upper panels) and five "noncategorical" subjects (lower panels): Percent "different" responses to pairings of a standard (S, stimulus No. 4) with seven comparison (C) stimuli, in four different vocalic context conditions.

frequent when the standard was paired with itself, and they increased with the physical distance of the comparison stimulus from the standard, with nearly perfect performance when the difference was 3 steps. This effect of step size was highly significant in an analysis of variance on physically different pairs only, $F(2,8) = 47.9$, $p < .001$. Due to the small number of subjects, no other effect reached conventional levels of significance. Nevertheless, the figure suggests two further effects: an increase in "different" responses when the periodic portions were different, $F(1,4) = 6.4$, $p < .10$, and a shift of the [(ʃ)ɑ]/[(s)u] discrimination function relative to the [(s)u]/[(ʃ)ɑ] function (right-hand panel).[3] Even though this latter effect did not approach statistical significance, it is of great interest that the shift occurred in a direction *opposite* to that exhibited by the categorical subjects (top right-hand panel). Inspection of individual subject data suggested that three listeners exhibited such a shift; the remaining two seemed to be unaffected by the nature of the periodic portion. Thus, although the data are not quite strong enough to warrant the conclusion that some of these listeners were indeed affected by the periodic stimulus portion, it is clear that they were *not* affected in the way the first group of subjects was.

GENERAL DISCUSSION

Summary of Results

The present study has three major results:

(1) Fricative identification is influenced by the periodic portion following the fricative noise, even when this portion is held constant over a block of trials. There are independent effects of formant transitions and vowel quality. This replicates the earlier findings of Whalen (in press) and Mann and Repp (1980). The effects were smaller here than in these earlier studies, but this reduction in size may have been due to the use of an improved [ʃ]-[s] continuum as well as to the blocked presentation of stimuli.

(2) Most naive subjects perceive fricative-vowel stimuli rather categorically, and they do so even in a fixed-standard AX task which was thought to provide a better opportunity for making auditory judgments. This result confirms the earlier findings of Fujisaki and Kawashima (Note 2) and of May (1979). While individual listeners may have varied somewhat in their ability to detect auditory differences between the stimuli, their judgments reflected primarily the phonetic category membership of the stimuli.

(3) Experienced listeners and some naive listeners were able to discriminate differences in fricative noise spectrum accurately and with little regard to the following periodic portion. If the periodic portion had any influence on their responses, it was in the opposite direction of the influence it had on categorical listeners. (We may disregard the bias to respond "different" when the irrelevant portions of the stimuli in a pair were different, which was perhaps shared by all listeners.) It is important to note that noncategorical listeners were not distinguished from categorical listeners in an identification task; all of them (whether experienced or not) showed the expected shifts in labeling functions contingent on vowel quality and formant transitions. (In the case of the experienced listeners, this fact was known from earlier studies.)

Two Listening Strategies

Obviously, the noncategorical subjects used a different listening strategy than the categorical subjects. That strategy was the one demanded by the instructions, viz., to focus attention on the auditory (essentially pitch-like) quality of the noise portion. Introspections and comments of the experienced listeners suggested that this strategy entailed a perceptual segregation of the noise portion from the periodic portion--a phenomenon related to auditory streaming (Bregman, 1978; Cole & Scott, 1973). Whether or not phonetic categorization is bypassed in the process, either deliberately or because the noise segregation prevents it, is not known. The author's experience as a listener suggests that some effort and attention are required to maintain a nonphonetic listening mode; however, another experienced listener commented that she easily and naturally segregated the noise portions. (The same listener shows large effects of vocalic context in an identification task; thus, she is able to integrate the two stimulus portions just as easily when the task requires it.)

That a nonphonetic strategy requires effort and, perhaps, some experience is also suggested by the performance of the categorical listeners. These subjects, even though they had been carefully instructed that subtle differences would occur in the noise portion alone, were apparently not able to follow the instructions effectively. It is a moot point whether an inferior ability to make fine auditory discriminations forced these listeners to remain in a phonetic mode, or whether their ability to focus attention on auditory properties of speech stimuli was less developed. However, the second possibility is far more plausible. After all, conscious access to auditory qualities of speech, particularly of those relatively brief segments that support phonetic perception, is rarely required of the ordinary speaker/hearer and has traditionally been the exclusive domain of phoneticians and speech scientists. Therefore, it should not be surprising that most naive listeners are not immediately able to perform this feat and instead show a strong tendency to persist in their habitual mode of phonetic perception. If their categorical behavior, especially in the fixed-standard AX task, was nevertheless a bit unexpected, it was only because fricative-vowel stimuli seem to offer a relatively easy opportunity to gain access to auditory stimulus properties. The noise portion is relatively steady-state and lasts 100-200 msec; no training is required for accurate detection of spectral differences when the portion occurs in isolation. Presumably, little training would be required to transform the catgorical listeners of the present study into noncategorical listeners, in contrast to the considerable training that is necessary for subjects to be able to discriminate fine differences in formant transitions or voice onset time of stop consonants (cf. Edman, 1979; Carney et al., 1977). In fact, the ability to focus attention on the noise portion of fricative-vowel stimuli might be discovered rather than learned, as suggested by the extremely accurate performance of two naive listeners. (One of them actually outperformed the three expert listeners.) However, this conjecture needs to be proven by further research.

Level of Vocalic Context Effects

The fact that noncategorical listeners were not significantly influenced by vocalic context indicates that effects of such context on fricative perception occur at a level that is sensitive to a listener's strategies. Since relatively low-level auditory phenomena--such as auditory masking or contrast--would seem less likely to depend on listening strategies, it is tempting to conclude that the effects of vocalic context are not of this class. However, it may be argued that too little is known about the influence of subjective perceptual organization on auditory interference and contrast, and that the differences between the present two groups of subjects may have resulted from different auditory strategies. Different subjects may have centered their attention on different parts of the signal: The noncategorical subjects may have paid attention to the onset of the fricative noise, where auditory interactions with the periodic portion were absent, whereas the categorical subjects may have focused on the offset of the fricative noise, where it adjoins the periodic portion and is most susceptible to auditory interference.[4] However, this argument should not distract from the fact that no convincing auditory explanation for the effects of vocalic context on fricative identification has yet been proposed. Likewise, there is no good auditory rationale for why listeners should vary in their perceptual strategies as they do, and it is not clear why paying attention to the initial portion of a fricative noise should lead to so much better discrimination performance than paying attention to its final portion.

On the other hand, there are numerous studies in the literature that suggest a phonetic origin for various contextual effects in speech perception (e.g., Bailey & Summerfield, 1980; Fitch, Halwes, Erickson, & Liberman, 1980; Mann, in press; Mann & Repp, 1980, in press; Repp, Liberman, Eccardt, & Pesetsky, 1978). Several other studies provide direct support for the existence of two distinct modes of processing speechlike stimuli--one auditory, the other phonetic (e.g., Bailey, Summerfield, & Dorman, 1977; Remez, Rubin, & Pisoni, in press; Grunke & Pisoni, Note 3). The strongest evidence on both counts comes from a recent study by Best, Morrongiello, and Robson (in press) who showed one type of cue integration (viz., integration of silence and formant transitions as cues to stop manner) to be specific to a phonetic mode of perception. Methodologically, the present study is complementary to that of Best et al.: Whereas they showed that certain (speechlike) nonspeech stimuli can be perceived either in an auditory or phonetic mode, the present experiments showed that the same is true for certain speech stimuli. In each case, the contextual or cue-integration effect of interest was observed only when listeners responded to phonetic, rather than auditory, properties of the stimuli.

We have noted that some of the noncategorical listeners appeared to be influenced by vocalic context, but in a direction opposite to that exhibited by the categorical listeners (and by themselves in an identification task). To the extent that these effects were real (and they could not be supported statistically), there are two possible explanations: (1) They may represent real auditory effects of the periodic portion on perception of the fricative noise. In this case, the effects of vocalic context observed in phonetic classification must have been phonetic in nature, as they overrode auditory effects of opposite sign. (2) Alternatively, some of the noncategorical listeners perhaps could not avoid classifying the stimuli into phonetic categories while, at the same time, they were judging the auditory quality of the noise portion. Since phonetic classification was probably influenced by vocalic context in the expected direction, it may have led to compensatory adjustments in the auditory judgments; e.g., an ambiguous noise categorized as "s" in [(s)u] context might have seemed unusually low-pitched for an "s". This explanation assumes that an auditory listening strategy does not preclude simultaneous phonetic categorization--an assumption that needs further testing.

Conclusion

The present data provide support for the hypothesis that effects of vocalic context on fricative identification are tied to a phonetic mode of perception. They suggest strongly that there are two different strategies of listening to fricative-vowel syllables, one auditory (noncategorical) and the other phonetic (categorical). Regular vocalic context effects occur only in the phonetic mode, presumably because they are mediated by the listener's implicit knowledge of articulatory patterns. Clearly, fricative-vowel syllables represent a category of speech sounds whose perception is neither categorical nor continuous but can be one or the other depending on listener strategy. Even though this is probably true for all speech sounds, fricative-vowel syllables differ from, say, stop-consonant-vowel syllables in that some of their auditory properties are easier to access. In summary, the present results reaffirm the importance of the distinction between auditory and phonetic perception, and they demonstrate that certain integrative processes are specific to the phonetic mode.

REFERENCE NOTES

1. Kunisaki, O., & Fujisaki, H. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics* (University of Tokyo), 1977, *11*, 85-91.
2. Fujisaki, H., & Kawashima, T. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute* (University of Tokyo), 1969, *28*, 67-73.
3. Grunke, M. E., & Pisoni, D. B. Some experiments on perceptual learning of mirror-image acoustic patterns. *Research on Speech Perception* (Dept. of Psychology, Indiana University), 1979, *5*, 147-182.

REFERENCES

Bailey, P. J., & Summerfield, Q. Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 536-563.

Bailey, P. J., Summerfield, Q., & Dorman, M. On the identification of sine-wave analogues of certain speech sounds. *Haskins Laboratories Status Report on Speech Research*, 1977, *SR-51/52*, 1-25.

Best, C. T., Morrongiello, B., & Robson, R. The perceptual equivalence of two acoustic cues for a speech contrast is specific to phonetic perception. *Perception & Psychophysics*, in press.

Bregman, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and Performance VII.* Hillsdale, NJ: LEA, 1978.

Carney, A. E., Widin, G. P., & Viemeister, N. F. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, *62*, 961-970.

Cole, R. A., & Scott, B. Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology*, 1973, *27*, 441-449.

Creelman, C. D., & Macmillan, N. A. Auditory phase and frequency discrimination: A comparison of nine procedures. *Journal of Experimental Psychology*, 1979, *5*, 146-156.

Edman, T. R. Discrimination of intraphonemic differences along two place of articulation continua. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers*. New York: Acoustical Society of America, 1979.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, 1980, *27*, 343-350.

Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1958, *1*, 1-7.

Hasegawa, A. *Some perceptual consequences of fricative coarticulation*. Unpublished doctoral dissertation, Purdue University, 1976.

Healy, A. F., & Repp, B. H. Context sensitivity and phonetic mediation in categorical perception: A comparison of four stimulus continua. *Haskins Laboratories Status Report on Speech Research*, 1980, *SR-63/64*, this volume.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, *54*, 358-368.

Mann, V. A. Influence of preceding liquid on stop consonant perception. *Perception & Psychophysics*, in press.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 1980, *28*, 213-228.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, in press.

May, J. G. *The perception of Egyptian Arabic fricatives*. Unpublished doctoral dissertation, University of Connecticut, 1979.

Mills, C. B. Effects of context on reaction time to phonemes. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 75-83.

Pollack, I., & Pisoni, D. B. On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, 1971, *24*, 299-300.

Remez, R. E., Rubin, P. E., & Pisoni, D. B. Speech perception without traditional speech cues. *Science*, in press.

Repp, B. H. Accessing phonetic information during perceptual integration of temporally distributed cues. *Journal of Phonetics*, 1980, *8*, 185-194.

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 129-145.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 621-637.

Swinney, D. A., & Prather, P. Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts. Perception & Psychophysics, 1980, 27, 104-110.

Whalen, D. H. Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. Journal of the Acoustical Society of America, in press.

FOOTNOTES

[1] Only three previous studies seem to have used fricatives in a categorical-perception paradigm, and none of them has been fully published. Fujisaki and Kawashima (Note 2) found better-than-chance within-category discrimination of stimuli from a [ʃe]-[se] continuum, but there was a marked peak in the discrimination function at the category boundary. The listeners in this Japanese study perceived fricative-vowel syllables only slightly less categorically than stop-consonant-vowel syllables. This result was replicated with Egyptian listeners by May (1979) who used an [əʃə]-[əsə] continuum. Hasegawa (1976) presented American listeners with a synthetic [ʃ]-[s] noise continuum in two different vocalic contexts, [i-] and [u-]. After demonstrating a shift in fricative labeling contingent on the preceding vowel, he found only rather weak evidence for discrimination peaks in the vicinity of the category boundary. Discrimination performance within phonetic categories was quite good, leading to the conclusion that the stimuli were not categorically perceived. However, the listeners in that study had some practice in the task; note also that, in contrast to the other studies, the fricatives were in syllable-final position, which may have enhanced auditory memory and thus facilitated discrimination.

[2] In piloting the AXB tapes, the author found that he, too, could discriminate the noises on every single trial. The 2-step comparisons were nevertheless chosen, since inexperienced listeners were expected to be less accurate.

[3] When the same data are converted into d' scores, it becomes evident that, despite the higher percentage of "different" responses in the conditions with different periodic portions, performance was actually somewhat poorer than in the conditions with identical periodic portions. Presumably, listeners altered their response criteria contingent on the relationship between the irrelevant stimulus portions (cf. the different false-alarm rates evident in Figure 3).

[4] More direct evidence on that point could be obtained in a reaction time task that varies noise durations, the prediction being that categorical listeners will be slowed down by an increase in noise duration while noncategorical listeners will be unaffected. In an earlier reaction-time study (Repp, 1980), I showed that naive listeners tend to wait for the opening (CV) transitions of intervocalic stops before making a phonetic decision, whereas experienced listeners can reach an early decision after hearing the closing (VC) transitions. The findings that an increase in fricative noise duration does not reduce the influence of following vocalic context on fricative labeling (Mann & Repp, 1980: Exp. 1) and that, in a phoneme monitoring task, reaction times to /s/ are longer than to /b/ (Mills, 1980;

Swinney & Prather, 1980) indeed suggest that listeners normally wait for the end of the noise and the onset of the periodic portion before deciding on the phonetic category of a fricative. However, if attention is restricted to the auditory quality of the noise portion, rather than to the phonetic category of the stimulus, such a waiting period becomes unnecessary.

CONTEXT SENSITIVITY AND PHONETIC MEDIATION IN CATEGORICAL PERCEPTION: A COMPARISON OF FOUR STIMULUS CONTINUA

Alice F. Healy and Bruno H. Repp

Abstract. Categorical perception is an ideal rarely, if ever, observed in the laboratory. Two separate requirements must be met for categorical perception: (1) predictability of discrimination performance from labeling performance, and (2) independence of labeling responses from stimulus context. In order to determine the extent to which instances of noncategorical perception are due to failures to meet one or both of these requirements, we employed four stimulus continua in AX discrimination and labeling tasks: stop-consonant-vowel (CV) syllables, steady-state vowels, fricative noises, and complex tones varying in timbre. We found that CV syllables departed from the ideal only because of contextual influences on labeling. Neither requirement was met by vowels or fricative noises, but fricative noises were less predictable than vowels, and vowels were somewhat less context independent than fricative noises. Surprisingly, the timbre stimuli were more predictable and showed smaller context effects than vowels or fricative noises. This finding was attributed to the shorter duration of the timbre stimuli, which may have prevented stable auditory memory traces.

INTRODUCTION

Categorical perception is a mode of perception in which stimuli are encoded in terms of a few discrete categories rather than in terms of continuous attributes. It is said to obtain when stimuli drawn from a physical continuum are discriminated not much better than would be predicted from a knowledge of the way in which they were assigned category labels. The degree of categorical perception of a stimulus set has typically been assessed by comparing results of a discrimination task with predictions derived from an independent identification task. However, Repp, Healy, and Crowder (1979)

Acknowledgment: This research was supported by the National Institute for Child Health and Human Development Grant HD01994 and Biomedical Research Support Grant RR05596 to Haskins Laboratories, and National Science Foundation Grants BNS77-00077 and BNS80-00283 to Yale University. Alice F. Healy was supported by a Senior Faculty Fellowship from Yale University during the final stages of preparing this manuscript. We are indebted to Robert G. Crowder, William K. Estes, Alvin M. Liberman, and Michael Studdert-Kennedy for helpful discussions about this research, and to Patti Jo Price and Mary LaRue for help with data analysis.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

pointed out that this method confounds two aspects of categorical perception: "context independence" (which they called "absoluteness") and "predictability". Context independence refers to the degree to which the phonetic categorization of a given stimulus is independent of the context in which it occurs. Predictability is the degree to which discrimination appears to be based on category labels, rather than on continuous sensory stimulus attributes. While a set of stimuli that is categorically perceived must satisfy both of these criteria, a set that is perceived not so categorically may be less context independent, less predictable, or both. In other words, subjects may change their (covert) labeling responses in the context of the discrimination task but nevertheless base their discrimination judgments on these labels; or it may be that discrimination is not based on category labels, whether or not they change as a function of context.

The acknowledgment that categorical perception involves two separate aspects that are confounded in the standard predictability test was originally made by Lane (1965) but subsequently rejected by Studdert-Kennedy, Liberman, Harris, and Cooper (1970) on the grounds that the standard test is sufficient to determine whether a stimulus continuum is categorically perceived. However, such a test cannot reveal the reasons for any deviations from the ideal pattern, and since deviations are almost always observed, their explanation is a central issue.

In their recent study, Repp et al. (1979) applied this logic to isolated vowels, a type of stimulus that has been shown by conventional methods to be perceived in a noncategorical fashion (e.g., Fry, Abramson, Eimas, & Liberman, 1962; Stevens, Liberman, Studdert-Kennedy, & Öhman, 1969). The stimuli used by Repp et al. formed an /i-I-ɛ/ continuum. Degree of context independence was assessed by examining whether the labeling of these vowels changed when they were paired with other vowels from the same continuum. Extent of predictability was determined by comparing the probabilities of assigning two vowels in a pair same or different phonetic labels to the probabilities of assigning "same" and "different" responses to precisely the same vowel pairs in a discrimination test. In addition, a standard single-item identification test was run. This methodology revealed that the presumed noncategorical perception of isolated vowels derived primarily from the context sensitivity of these stimuli: Once context-induced (invariably contrastive) shifts in labeling probabilities were taken into account, discrimination performance could be predicted fairly closely, thus leaving open the possibility that vowel discrimination is mediated in large part by phonetic categories.

This result suggested to us that context sensitivity and phonetic mediation (predictability) are independent aspects of perception. Repp et al. (1979) hypothesized (in their "all-phonetic model") that contextual influences arise prior to categorization via a mechanism of auditory contrast similar to lateral inhibition, while the predictability of discrimination performance reflects the listeners' reliance on category labels and their reluctance or failure to refer to additional auditory stimulus information. According to that view, the size of context effects is determined by auditory stimulus properties, whereas the extent to which discrimination can be predicted from labeling presumably depends both on the relative accessibility of auditory stimulus information (cf. Fujisaki & Kawashima, 1969) and on the familiarity of the categories used. If contextual influences are relatively independent of the use of category labels in discrimination, then it might be possible to find a stimulus set that, unlike isolated vowels, shows small context effects (i.e., context independence) but poor predictability. In addition, of course, there may be stimulus sets that are high or low on both of these dimensions.

EXPERIMENT

In the present study, we compared four different stimulus sets with regard to the context independence and predictability criteria, using the methodology of Repp et al. (1979). We expected these stimulus sets to exhibit quite different patterns of results, as explained in more detail below. Thus, the results of our experiment were expected to bear on the question of whether context independence and predictability are independent aspects of categorical perception.

Our first set of stimuli was a continuum of CV syllables ranging from /ba/ to /da/. It is well known that these stimuli are perceived highly categorically (e.g., Liberman, Harris, Hoffman, & Griffith, 1957). Therefore, they were expected to be high on both the context independence and predictability criteria. Nevertheless, there was more to be learned about their perception. We were interested in whether they show any reliable context effects at all, and if so (cf. Eimas, 1963; Rosen, 1979), how the magnitude of these effects compares to those found for other stimuli. It is a common finding in conventional studies of categorical perception that discrimination performance is somewhat higher than predicted, even for stimuli that are perceived highly categorically. We wondered whether this discrepancy could be accounted for by context effects in covert labeling; perhaps, the difference would disappear when "in-context" predictions (derived from subjects' labeling responses to stimuli presented in the same format as in the discrimination task) are used.

Our second set of stimuli was a continuum of isolated vowels ranging from /i/ to /I/. This part of the experiment was expected to provide a partial replication of the Repp et al. results and a basis for a more direct comparison with the other stimulus sets. On the basis of the Repp et al. findings, we expected the vowels to exhibit large contrast effects in labeling but relatively high predictability of discrimination scores from in-context labeling results. Whether predictability would be as high for vowels as for CV syllables was of particular interest, because of the suggestion by Repp et al. (1979) that vowels may be as predictable as CVs.

Our third set of stimuli was a continuum of isolated fricative noises ranging from /ʃ/ to /s/. Considerably less was known about the perception of these stimuli than about the preceding two sets. However, Mann and Repp (1980) recently used them in several labeling tasks and found that subjects assigned them to phonetic categories reliably and without difficulty. Informal observations also suggested that these noises were not particularly sensitive to context and easy to discriminate. Thus, this stimulus set was a candidate for being high on context independence but low on predictability--a result that would indicate that the two dimensions can be dissociated. This part of the experiment also served as a partial replication of a previous study by Fujisaki and Kawashima (1969) who--to the best of our knowledge--were the only authors ever to use a continuum of isolated fricative noises in a categorical perception task. They, like Mann and Repp (1980), found very reliable identification of these noises, as well as better-than-chance discrimination within phonetic categories. However, they also found a marked discrimination peak at the category boundary--a finding that was taken to indicate the involvement of phonetic categories in discrimination. We wondered whether this result could be replicated.

Our fourth set of stimuli was a continuum of brief complex tones varying in timbre. They were isolated synthetic single-formant resonances varying in frequency, but with a constant fundamental frequency. The categories subjects used in classifying these stimuli were "low" and "high," referring to their relative pitch ("dull" and "sharp" or "dark" and "bright" might have been equally appropriate labels). Although this stimulus continuum had some aspects in common with a vowel continuum, it was expected to be perceived noncategorically, like other physical continua of simple nonspeech sounds. Classification into essentially arbitrary categories was expected to be highly context-dependent, and predictability was expected to be poor, because of the absence of mediation by category labels.

Each of the four stimulus continua had the same number of stimuli (10) and categories (2). Since it is difficult to equate relative discriminability across continua without extensive pilot work, we instead chose to present stimulus comparisons one, two, and three steps apart on each continuum. Thus, one-step differences on a continuum of easily discriminable stimuli might give performance levels comparable to those of two-step or even three-step differences of other stimuli that were more difficult to tell apart.

Aside from its primary purpose--the separation of the two aspects of categorical perception--our study served as a detailed investigation of perceptual contrast effects, i.e., the tendency to give successive stimuli different labels. We were in a position not only to compare the magnitudes of contrast effects across different stimulus continua but also to compare forward and backward contrast effects within stimulus pairs, and to investigate the influence of varying step size (i.e., physical stimulus difference) on the size of contrast. We hoped that our results would bring us closer to an understanding of the stimulus characteristics that facilitate or inhibit contrast between successive stimuli.

Method

Subjects. The subjects were 12 paid volunteers, men and women recruited by posters on the Yale University campus. None of them was experienced in discrimination tasks, although several had listened to synthetic speech for other experimental tasks conducted in our laboratory.

Stimuli. Four different continua of synthetic sounds were used. Each continuum contained 10 stimuli spaced in approximately equal physical steps. The first three (speechlike) continua were generated on the OVE IIIc serial resonance synthesizer at Haskins Laboratories; the fourth (nonspeech) continuum was created on the Haskins Laboratories parallel resonance synthesizer.

The *CV syllables* (/ba/-/da/) differed in the onset frequencies of the second and third formants, which are listed in Table 1. The transitions from these onset frequencies to the formant steady-states (at 1233 and 2520 Hz, respectively) were stepwise-linear and 40 msec in duration. All CV syllables had in common a 30-msec transition in the first formant (from 200 to 771 Hz), a fundamental frequency contour that was steady at 125 Hz over the first 50 msec and then fell linearly to 80 Hz, a flat amplitude contour with a final ramp, and a total duration of 250 msec.

Table 1

Stimulus Parameters (in Hz)

Stim.	CV Syllables		Vowels			Fric. Noises		Timbres
No.	F2	F3	F1	F2	F3	P1	P2	(F2)
1	859	1795	269	2296	3019	1957	3803	2156
2	937	1929	281	2263	2976	2197	3915	2234
3	1022	2059	293	2247	2933	2466	4148	2307
4	1099	2197	304	2214	2912	2690	4269	2387
5	1181	2328	315	2198	2870	2933	4394	2462
6	1260	2466	327	2167	2829	3199	4655	2540
7	1345	2594	339	2151	2789	3389	4792	2615
8	1425	2729	351	2120	2749	3591	4932	2692
9	1510	2870	364	2105	2709	3917	5077	2762
10	1588	2998	375	2075	2670	4243	5322	2837

The *vowels* (/i/-/I/) differed in the frequencies of the first three formants, which are listed in Table 1. All vowels were completely steady-state, with a linearly falling fundamental frequency contour (from 125 to 80 Hz), a flat amplitude contour with initial and final ramps, and a total duration of 250 msec. Due to synthesizer characteristics, stimulus amplitude increased slightly across the continuum.

The *fricative noises* (/ʃ/-/s/) differed in the frequencies of two fricative formants (poles), which are listed in Table 1. All stimuli were steady-state, had flat amplitudes with initial and final ramps, and a total duration of 250 msec. Due to certain adjustments in the amplitude specifications at the synthesis stage, the stimuli had increasingly lower amplitudes (a total decrease of about 4 dB), flatter amplitude ramps, and relatively more abrupt onsets towards the high (/s/) end of the continuum. These factors may have contributed to the discriminability, of the noises, but this contribution was expected to be small because differences in noise spectra were quite salient to begin with.

The timbres ("low"-"high") were single (second-)formant resonances varying in frequency (see Table 1). All timbres were steady-state, with a fundamental frequency of 124 Hz, a flat amplitude contour, and a total duration of 50 msec. The short duration was chosen to reduce the speechlikeness of the stimuli (250-msec timbres sounded vowel-like) as well as their discriminability, which seemed too high initially. (Spacing on the continuum could not be reduced because of synthesizer limitations.)

For each of the four stimulus sets, two tapes were recorded using the Haskins Laboratories stimulus sequencing program. Except for the differences in stimuli, these tapes were identical for all four sets. The simple identification tapes contained 20 repetitions of each of the 10 stimuli on a given continuum, arranged in 4 random sequences of 50 (5 repetitions of each stimulus) with 3-sec interstimulus intervals (ISIs). In addition, the two endpoint stimuli of the continuum were recorded five times in alternation at the beginning of the tape, to provide examples of the two categories. The AX tapes contained 4 random sequences of 68 stimulus pairs, with 300-msec ISIs within pairs and 4-sec ISIs between pairs. The 68 pairs in a block included the 10 identical, 9 one-step, 8 two-step, and 7 three-step pairs, in both possible stimulus orders [2 x (10 + 9 + 8 + 7) = 68].

Procedure. Each subject participated in four sessions, one for each stimulus type. The sequence of stimulus types was counterbalanced across subjects according to a Latin square design. There were three tasks in each session; the sequence of tasks was likewise counterbalanced across subjects but was fixed for a given subject across the four sessions.

In the simple identification task, the subjects were first presented with the alternating endpoint stimuli to exemplify the response categories. Then, they assigned in writing a label to each stimulus heard. The symbols used for the four stimulus types were: b, d (CV syllables); i, I (vowels); sh, s (fricative noises); L, H (timbres).

In the AX labeling task, the subjects assigned labels to both stimuli in each pair. The same labels were employed as in the simple identification task. If the AX labeling task was first in a session, it was preceded by examples of the endpoint stimuli (from the simple identification tape). In the AX discrimination task, only the responses changed; they were now s (same) and d (different), and the subjects were carefully instructed to listen for any difference between the stimuli. In all conditions, the subjects were given a brief preview of the tapes before responding began: A randomly selected section was played for 1-2 minutes, and subjects listened without responding.

The subjects listened to the stimulus tapes in a quiet room over TDH-39 earphones. The tapes were played back on an Ampex AG-500 tape recorder at a comfortable loudness. Due to their different acoustic characteristics, the different stimulus types varied somewhat in overall amplitude, but all were within a comfortable listening range.

Results and Discussion

Simple identification. The results of the single-item identification test are summarized in Figure 1 in terms of percentages of "b" and "d" responses for CV syllables, "i" and "I" responses for vowels, "sh" and "s" responses for fricative noises, and "L" and "H" responses for timbres. The CV syllables differ from the other three stimulus sets in that the labeling functions are steeper and the category boundary (the 50-percent cross-over point of the labeling function) is definitely off-center (the "b" category being larger than the "d" category), whereas the other category boundaries fall close to the centers of the respective continua (between stimuli 5 and 6). This pattern of results, which was also found at the individual level, indicates a certain amount of context independence of CV syllables. The arbitrary category boundary for timbres was naturally expected to fall right in the center, as it did; the central locations of the vowel and fricative boundaries may have been simply a consequence of our selection of stimulus ranges.

We also used these identification results to predict discrimination performance, following the classical "low-threshold" model (Pollack & Pisoni, 1971). The resulting predictions, averaged over subjects, are represented in the top row of Figure 2 in terms of percent "different" responses as a function of stimulus number and step size.

Predictability. The results of the AX discrimination task are displayed in the bottom row of Figure 2 in terms of percent "different" responses as a function of stimulus number and step size. In the center row of Figure 2 are the corresponding scores ("in-context" predictions) derived from the AX labeling task by computing the percentages of trials on which the two stimuli in a pair were given different labels.

Separate analyses of variance for each step size of each stimulus type were performed to compare the discrimination functions to the analogous functions based on AX labeling. These analyses revealed a significant discrepancy in favor of the discrimination task for each stimulus type at each step size ($p < .05$ or less in each case). However, these significant differences between tasks do not in themselves imply that performance was significantly better than in the discrimination task, since both hits (1- to 3-step functions) and false alarms (0-step functions) showed larger values than in the labeling task, indicating that subjects had a greater tendency to respond "different" in the discrimination task (particularly with CV syllables and timbres). In order to control for this response bias, values of d' were obtained from the tables provided for the AX paradigm by Kaplan, Macmillan, and Creelman (1978). To obtain relatively stable estimates of d', it was necessary to average hit rates (separately for the three step sizes) and false alarm rates (based on pairs of identical stimuli) across stimulus pairs on each continuum before determining d' values for each subject and each stimulus type.[1] The values of d', averaged across subjects, are shown in Table 2.

An analysis of variance of these d' values included the following factors: step size, task (discrimination vs. labeling), and stimulus type. The overall difference between discrimination and labeling tasks was significant, $F(1,11) = 60.8$, $p < .001$, as was the interaction of stimulus type and task, $F(3,33) = 48.0$, $p < .001$. The performance level in the discrimination

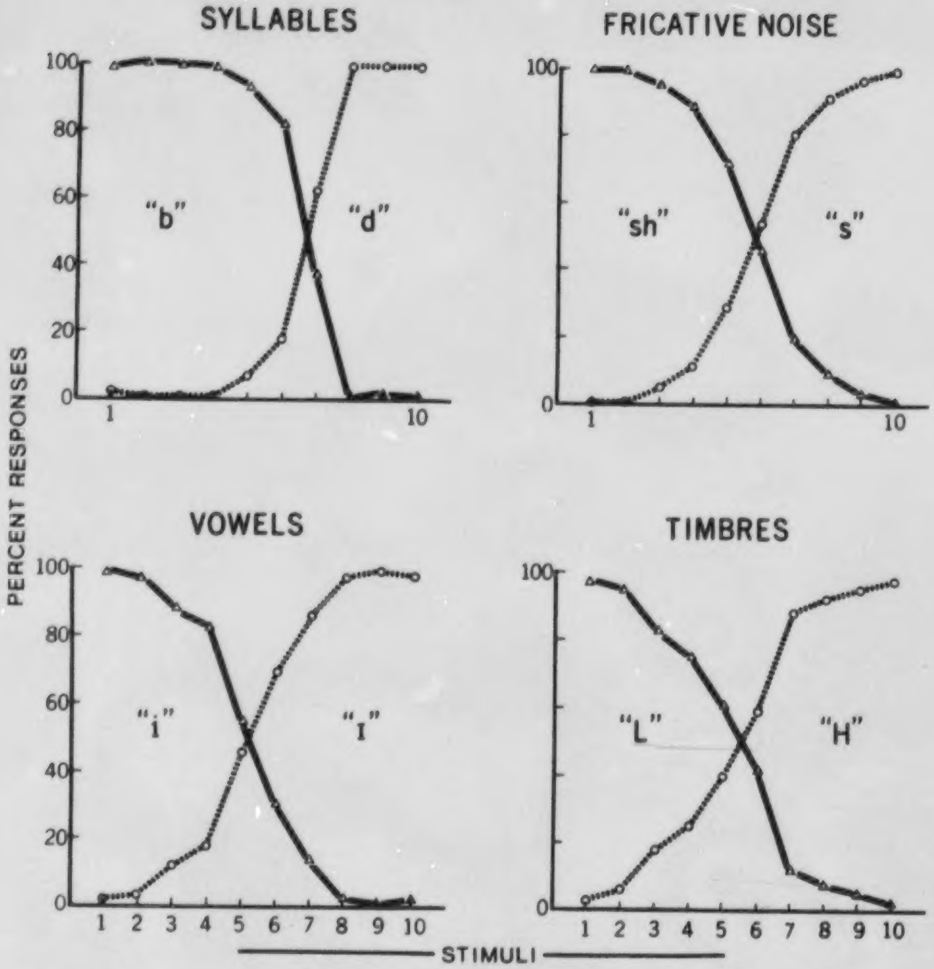


Figure 1. Labeling functions for the four stimulus continua in the simple identification task.

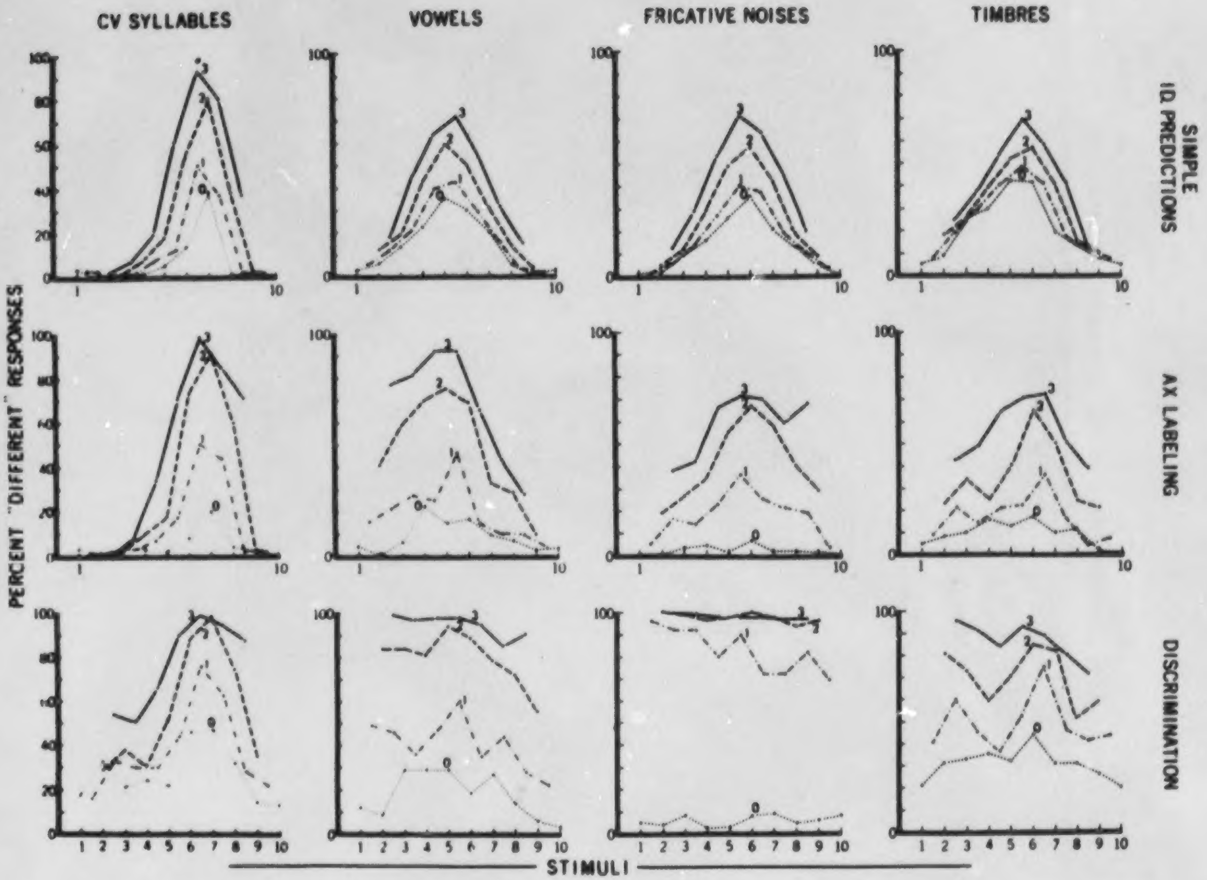


Figure 2. Percent "different" responses in the AX discrimination task (bottom row), in the AX labeling task (middle row), and as predicted from simple identification (top row).

task exceeded that in the AX labeling task for timbres, F(1,11) = 7.5, p = .019, for vowels, F(1,11) = 21.4, p = .001, and especially for fricative noises, F(1,11) = 131.8, p < .001, whereas AX labeling performance actually exceeded discrimination performance for CV syllables, although only with marginal significance, F(1,11) = 4.5, p = .056. The reversal for CV syllables suggests that listeners, in their (unsuccessful) attempt to make fine discriminations among CV syllables, made less effective use of category labels than in the labeling task. It also suggests that the commonly observed advantage of obtained CV syllable discrimination over scores predicted from single-item identification tests may indeed be due to context effects in the discrimination paradigm (see below)--i.e., that the advantage is an artifact of using inappropriate predictions. For vowels, the significant advantage of discrimination over labeling performance indicates that, contrary to the preliminary conclusions of Repp et al. (1979), the discrimination of isolated steady-state vowels is not phonetically mediated to the same extent as the discrimination of CV syllables. Phonetic mediation seems to play little or no role in fricative noise discrimination, where performance was exceedingly high even within categories.

Table 2

Average values of d'
as a function of task and step size for each stimulus type

		Step Size		
		1	2	3
CV Syllables				
	Labeling	1.20	2.14	2.90
	Discrim	0.93	1.75	2.90
	D-L	-0.27	-0.39	0.00
Vowels				
	Labeling	1.24	2.41	3.15
	Discrim	1.57	3.32	4.38
	D-L	0.33	0.91	1.23
Fricative Noises				
	Labeling	1.90	2.90	3.59
	Discrim	4.69	5.80	5.78
	D-L	2.79	2.90	2.19
Timbres				
	Labeling	0.82	1.75	2.54
	Discrim	1.30	2.36	3.39
	D-L	0.48	0.61	0.85

Clearly, the magnitude of the overall difference between discrimination and labeling performance cannot be taken as a direct indicator of whether or not discrimination responses are mediated by category labels. Even if category labels play no role, discrimination performance will approach labeling performance when discrimination is made sufficiently difficult. To assess the possible role of mediation by category labels, the shapes of the obtained discrimination and labeling functions need to be compared as well. If category labels were used in the discrimination task, performance should be better in the category boundary region than within categories. Thus, discrimination scores should show peaks at the same points as AX labeling scores. (Compare the figures in the bottom row with those in the middle row of Fig. 2.)

Such peaks are clearly present in the discrimination functions for CV syllables. The vowels show small peaks in the boundary region, especially in the 1-step function, indicating that category labels did play some role. Performance with fricative noises was too close to the ceiling, at least for 2- and 3-step functions, for any clear peaks to be exhibited. The timbre results are puzzling: The discrimination functions (especially 1-step and 2-step) do exhibit peaks in the category boundary region,[2] even though it might seem implausible that the subjects relied on the arbitrary category labels, "high" and "low," in making their discriminations. However, there is no obvious psychoacoustic reason why discriminability should have been higher in the center of the timbre continuum. We will return to this unexpected result with timbres in our discussion below. In summary, the question of whether mediation by category labels played a role in discrimination is to be answered as follows: CV syllables--yes; vowels--in part; fricative noises--can't tell (if yes, category labels had little to contribute); timbres--in part (surprisingly).

For three of the stimulus types--vowels, fricative noises, timbres--the listeners must have made (additional) use of auditory information in the discrimination task. Auditory information should become more available as the physical stimulus differences increase. As can be seen in Table 2, both labeling and discrimination d' scores increase with step size. However, to reflect a true increase in auditory information, discrimination scores should increase more than labeling scores--i.e., the difference between labeling and discrimination scores should increase as a function of step size. Such an increase can indeed be observed for vowels [the interaction of task and step size was significant, $F(2,22) = 9.5$, $p = .001$] and--to a much smaller extent--for timbres, $F(2,22) = 2.6$, $p = .097$. For fricative noises, the results were distorted by a ceiling effect; otherwise, they presumably would have shown a similar pattern. For CV syllables, the increase between step sizes 2 and 3 (Table 2) was not significant. This pattern of results further establishes that additional auditory information is available for vowels, timbres, and most likely fricative noises, but not for CV syllables.

Context independence. In order to assess the effects of stimulus context on identification in the AX labeling tasks of the present experiment, we tabulated the labeling response frequencies separately for stimuli occurring first and those occurring second in the stimulus pairs, and we then examined these frequencies for one (target) stimulus contingent on the nature of the other (nontarget) stimulus in the pair. Only target stimuli 4-7 were considered, since the other stimuli could not be paired with both higher and lower stimuli one, two, and three steps apart on a given continuum. The results are shown in Figure 3: The percentage of responses in the "lower"

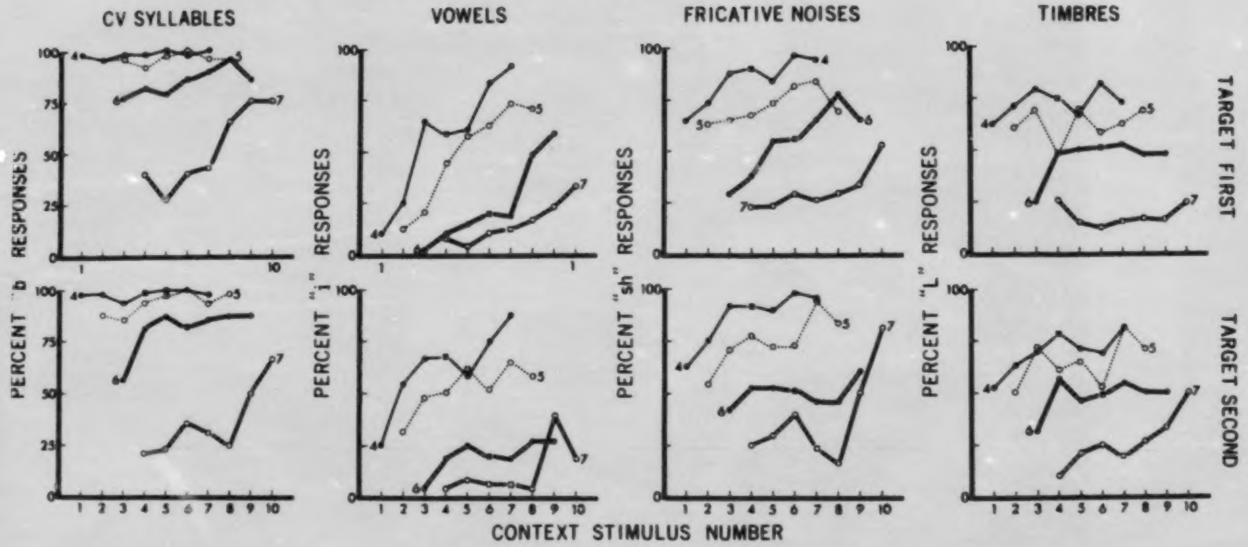


Figure 3. Context effects in the AX labeling task: Percent responses in the category associated with stimulus 1, plotted as a function of target stimulus position (first or second), target stimulus number, and context stimulus number. Pairs of identical stimuli are represented by squares.

response category (the category associated with stimulus 1) is shown, separately for each target stimulus, as a function of the identity of the context (nontarget) stimulus. Separate panels are provided for targets in first and second position. A contrast effect appears as a positive slope of the lines in each graph, whereas a flat function would imply no contrast.

It can be seen that all four stimulus types exhibit contrast effects: The percentage of responses in the "lower" category was greater when the context stimulus was above than when it was below the target on the continuum, $F(1,11) = 46,4$, $p < .001$.[3] However, the magnitude of the effect varies with stimulus type--the interaction of stimulus type and position of context stimulus relative to target (lower versus higher) was significant: $F(3,33) = 3.7$, $p = .022$. This interaction may be due in part to a ceiling effect for stimuli 4 and 5 of the CV syllables. Note that CV stimulus 7 shows contrast effects comparable in magnitude to those obtained with vowels. Separate analyses conducted on each stimulus type revealed significant contrast effects for vowels, $F(1,11) = 56.7$, $p < .001$, CV syllables, $F(1,11) = 39.2$, $p < .001$, and fricative noises, $F(1,11) = 10.2$, $p = .008$, but not for timbres, $F(1,11) = 2.3$, $p = .153$. In accordance with the data of Repp et al. (1979), retroactive contrast (target first) was significantly larger than proactive contrast (target second) for vowels, $F(1,11) = 8.5$, $p = .014$. None of the other stimulus types showed a significant difference in this direction; timbres actually showed a tendency in the opposite direction.

The percentage of responses in the "lower" category increased with context stimulus position on both sides of the target, $F(2,22) = 82.9$, $p < .001$. This increase was greater for some stimulus types than for others, as revealed in a significant interaction of context stimulus position and stimulus type, $F(6,66) = 4.7$, $p = .001$. This interaction may also be due in part to a ceiling effect for the CV syllables. Separate analyses conducted on each stimulus type revealed significant effects of context stimulus position for each [vowels: $F(2,22) = 53.8$, $p < .001$; CV syllables: $F(2,22) = 6.9$, $p = .005$; fricative noises: $F(2,22) = 28.8$, $p < .001$; timbres: $F(2,22) = 4.9$, $p = .017$].

According to these results, timbres are highest in context independence (quite unexpectedly), with considerable contrast effects for fricative noises, CV syllables, and especially vowels. Note that the context effects obtained for the various stimulus types do not always take the same form. For example, retroactive contrast effects are larger than proactive effects for vowels, but retroactive and proactive contrast effects are essentially equal for CV syllables. The effects of stimulus context therefore depend on the nature of the stimulus, and a simple explanation of these effects will not hold across different stimulus types.[4]

GENERAL DISCUSSION

"Categorical perception" is often understood to refer to the use of categories in discrimination (e.g., Macmillan, Kaplan, & Creelman, 1977); however, examination of the source literature (Liberman et al., 1957; Studdert-Kennedy et al., 1970) reveals that "categorical" was originally intended to mean "absolute." Thus, the original definition of categorical perception includes as criteria both context independence and the use of categories ("predictability"). One of the aims of the present study was to separate these two aspects, by examining to which extent different sets of stimuli satisfy one or the other. Our results show that the two aspects are at least partially independent: Stimuli may exhibit large contrast effects even though discrimination is partially based on category labels (as in the case of vowels), or they may be less sensitive to context even though category labels play little role in discrimination (as in the case of our fricative noises). Both vowels and fricative noises are noncategorically perceived, but apparently for different reasons--vowels primarily due to context sensitivity, fricative noises primarily due to lack of predictability.

Using the methodology proposed by Repp et al. (1979), we demonstrated that discrimination performance for CV syllables does not exceed labeling performance when context effects on labeling are taken into account (so-called "in-context" predictions). Thus, the small discrepancy between predicted and obtained discrimination performance in past studies was most likely due to context effects in covert labeling during the discrimination task. Our results strongly support the hypothesis that listeners, at least naive ones, discriminate CV syllables by relying exclusively on phonetic category information. In fact, the task requirement of detecting within-category distinctions seems to lead to a somewhat *less* efficient use of category labels, but not to the recovery of auditory information. However, it has been shown that auditory properties of stop consonants differing in place of articulation do become available after discrimination training (Edman, 1979).

A comparison of the results of vowels and fricative noises is revealing with regard to the possible determinants of context independence and predictability. In both stimulus types, the distinctive spectral properties were constant throughout the stimulus duration, which was the same for vowels and fricative noises, and the labeling functions for the two stimulus continua were quite similar. However, discrimination performance was much higher for fricative noises than for vowels. Discrimination performance for 2-step vowel pairs was similar to that for 1-step fricative noise pairs (cf. Figure 2), so a fair comparison can be made between those portions of the results. However, even when the obtained performance levels are thus equated, it is still true that vowels are more predictable (i.e., a larger portion of the discrimination scores can be accounted for by the use of category labels), whereas fricative noises are less context-sensitive. How are these differences to be explained?

The difference in predictability could arise from either or both of two sources: a difference in auditory distinctiveness, or a difference in the use of category labels in discrimination. The much higher discrimination scores for fricative noises may reflect the greater auditory distinctiveness of these stimuli; in addition, however, listeners may have been able to ignore category labels and thus to access auditory information more successfully with fricative noises than with vowels. In other words, the noises, being less speechlike, may have facilitated an auditory mode of processing.

The difference in the contrast effects exhibited by vowels and fricative noises is harder to explain. Although this difference is small overall, it is considerable when discrimination performance is equated (1-step fricative noises vs. 2-step vowels). Some investigators have argued that contrast effects arise only after categorization of the stimuli (Fujisaki & Shigeno, 1979), but there is evidence that this argument is not correct. Specifically, Repp et al. (1979) found that contrast effects were greatly diminished when an irrelevant sound was interpolated between the two sounds in an AX pair. Such a manipulation should affect auditory (or precategorical) memory but not phonetic (or categorical) memory. Therefore, we must look at the auditory properties of the stimuli in order to understand the basis for the contrast phenomenon. The primary difference in auditory terms between vowels and fricative noises seems to be the periodic versus aperiodic nature of the waveform. Perhaps it is with periodic stimuli such as vowels that especially large contrast effects are found. (See May, 1979, for a similar hypothesis.) Clearly, this hypothesis requires further testing (e.g., by using whispered vowels).

The pattern of results for the nonspeech stimuli, the timbres, was unexpected in several respects. We expected timbres to be the least categorically perceived of the stimuli we studied, since the category labels attached to the stimuli were completely relative. For that reason, it seemed unlikely that subjects would base their responses on the category labels or that the category labels would be stable across changes in stimulus context. On the contrary, we found a fair amount of predictability for timbres. In fact, the labeling performance for timbres matched the discrimination performance more closely than was the case for vowels (but less closely than for CV syllables). In addition, peaks at the category boundary region were found in the discrimination functions, although these peaks were considerably smaller than those found for CV syllables. Moreover, the magnitude of the context effects on labeling was smaller for timbres than for any of the other stimulus classes studied. Therefore, timbres tended to satisfy both of the criteria for categorical perception, despite their status as nonspeech sounds and despite the arbitrary character of their category labels.

In attempting to explain these unexpected results, we are inevitably led to consider the fact that the timbre stimuli were very short in duration. Whereas all the other stimuli employed were 250 msec long, the timbres were only 50 msec. This short duration was necessary in order to insure that our timbres would not be mistaken for vowels. Fujisaki & Kawashima (1969) and Pisoni (1973) have reported that short vowels are perceived more categorically than long vowels, presumably because they have a less stable representation in auditory memory, which increases listeners' reliance on category labels. Likewise, our subjects may have been forced to rely on category labels, albeit arbitrary ones, in discriminating the short-duration timbres, because they were unable to hold these sounds in auditory memory. This argument is consistent with the fact that the critical portion of the highly predictable CV syllables was quite short in duration, although the entire stimulus was 250 msec long.

An explanation must still be found for the fact that timbres were high in context independence as well as predictability. The short duration of the stimuli may have been critical in this regard as well, since stable auditory memory traces may be required for contrast effects to be exhibited. However, duration per se may not provide a sufficient account for the context effects obtained in this experiment. The fricative noises were as long in duration as the steady-state vowels but exhibited a smaller contrast effect. In addition, Fujisaki and Shigeno (1979) have reported relatively small contrast effects with timbres that were 100 msec in duration, whereas they found larger contrast effects for vowels of the same duration.

The relatively high auditory similarity of the timbre stimuli (as evidenced by their poor discriminability) may be another factor that contributed to the weakness of the contrast effect. Indeed, Fujisaki and Shigeno (1979) have demonstrated that the magnitude of the contrast effects is decreased when the stimuli being compared are highly similar. (See also Crowder, 1980, for a relevant discussion.) Our own data corroborate these findings, since we also found smaller contrast effects for pairs of stimuli that were adjacent to each other on the continuum. (Note the tendency for the functions in Figure 3 to be flatter in the vicinity of the squares representing the identical pairs.) However, this line of reasoning would lead one to expect the largest contrast effects with fricative noises, since they were discriminated most easily. Instead, the fricative noises showed contrast effects that were smaller than those for vowels. Hence, auditory similarity alone cannot account for the magnitude of the contrast effects obtained with a given set of stimuli.

In conclusion, stimulus continua rarely, if ever, perfectly satisfy the standard predictability test, in which discrimination performance is predicted from performance on a single-item identification test. We have focused on two important causes for these departures from the ideal: Either the subjects may not rely wholly on category labels in discrimination, or the labels they use may be subject to contextual influences. Our data suggest that these two factors may vary independently. In particular, we have shown that the departure from the ideal for CV syllables is due entirely to contextual influences on labeling. We have also shown that fricative noises and vowels are perceived noncategorically for both reasons, but with context effects playing a larger role for vowels and reliance on auditory information playing a larger role for fricative noises. The nonspeech continuum of timbres that we studied surprisingly proved to be more categorically perceived than either fricative noises or vowels, due both to smaller context effects and to greater apparent reliance on category labels, albeit arbitrary ones. We tentatively ascribe this finding to the short duration of these stimuli, which may have prohibited the development of stable auditory memory traces.

REFERENCES

Crowder, R. G. The role of auditory memory in speech perception and discrimination. *Haskins Laboratories Status Report on Speech Research*, 1980, *SR-62*, 187-204.

Edman, T. R. Discrimination of intraphonemic differences along two place of articulation continua. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th Meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

Eimas, P. D. The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, 1963, *6*, 206-217.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. The identification and discrimination of synthetic vowels. *Language and Speech*, 1962, *5*, 171-189.

Fujisaki, H., & Kawashima, T. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute*, University of Tokyo, 1969, *28*, 67-73.

Fujisaki, H., & Shigeno, S. Context effects in the categorization of speech and nonspeech stimuli. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers presented at the 97th Meeting of the Acoustical Society of America*. New York: Acoustical Society of America, 1979.

Kaplan, H. L., Macmillan, N. A., & Creelman, C. D. Tables of d' for variable-standard discrimination paradigms. *Behavior Research Methods & Instrumentation*, 1978, *10*, 796-813.

Lane, H. The motor theory of speech perception: A critical review. *Psychological Review*, 1965, *72*, 275-309.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, *54*, 358-368.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. The psychophysics of categorical perception. *Psychological Review*, 1977, *84*, 452-471.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 1980, *28*, 213-228.

May, J. G. *The perception of Egyptian Arabic fricatives*. Unpublished doctoral dissertation, University of Connecticut, 1979.

Pisoni, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, *13*, 253-260.

Pollack, I., & Pisoni, D. B. On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, 1971, *24*, 299-300.

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 129-145.

Rosen, S. M. Range and frequency effects in consonant categorization. *Journal of Phonetics*, 1979, *7*, 393-402.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Öhman, S. E. G. Cross-language study of vowel perception. *Language and Speech*, 1969, *12*, 1-23.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, 1970, *77*, 234-249.

FOOTNOTES

[1] Unequal frequencies of individual stimuli were taken into account, and values of 0 and 1 were treated as .01 and .99, respectively, in the table look-up (d'max = 6.93).

[2] Analyses of variance performed on the discrimination data yielded significant effects of stimulus location ($p < .01$) for each step size of the timbres.

[3] For the purpose of this analysis, responses to pairs of identical stimuli (indicated by squares in Figure 3) were not included.

[4] Another effect that also varied considerably across stimulus types was that of stimulus order. Although vowels did not show any consistent overall effect of stimulus order, the interactions of stimulus order and position were highly significant ($\underline{p}$ = .002 or less) at all three step sizes: At the left (/i/) end of the vowel continuum, more "different" responses were obtained in both discrimination and labeling tasks when the first stimulus in a pair had a higher position on the continuum than the second, but this effect was reversed at the right (/I/) end of the continuum. This stimulus order effect is similar to one found in the study by Repp et al. (1979), although the reversal occurs at an earlier point on the vowel continuum in the present study.

CV syllables showed stimulus order effects, but their direction was inconsistent across different step sizes. For fricative noises, the high performance level may have prevented strong order effects. Timbres, when arranged from high to low frequency—in analogy to the second formant of the vowel continuum, which was in the same frequency range--showed weak trends in the same direction as vowels. These differences in the nature and size of the stimulus order effects as a function of stimulus type imply that these effects are not artifacts of the experimental design but rather reflect properties of the stimuli employed.

BIDIRECTIONAL CONTRAST EFFECTS IN THE PERCEPTION OF VC-CV SEQUENCES

Bruno H. Repp

Abstract. The two stop consonants in VC_1C_2V sequences are not perceptually independent: There are perceptual interactions in both directions, which tend to be contrastive unless the closure interval between VC_1 and C_2V is very short. Backward contrast tends to be larger than forward contrast; it declines as the closure interval is increased and is strongly influenced by the range of closure durations employed, whereas forward contrast is quite insensitive to these factors. Significant contrast effects are also obtained in a discrimination task, which contradicts explanations based on response bias. It seems likely that the demonstrated effects arise from listeners' knowledge of articulatory/acoustic speech patterns, perhaps from a perceptual compensation for coarticulatory dependencies between stops produced in sequence.

INTRODUCTION

There is ample evidence that speech perception is not a simple left-to-right process in time. The perception of a phonetic segment often depends on the following as well as on the preceding context. For example, the perception of a fricative consonant is influenced by the following vowel (Kunisaki & Fujisaki, Note 1; Mann & Repp, 1980), whereas the perception of a stop consonant in a cluster is affected by the identity of a preceding liquid or fricative (Mann, in press; Mann & Repp, in press). Even more striking examples of such contextual effects, both forward and backward in time, are provided by demonstrations that the perception of a syllable-final stop may depend on the duration of a fricative noise in the next syllable (Repp, Liberman, Eccardt, & Pesetsky, 1978) or on the nature of the initial consonant of the same syllable (Raphael, Dorman, & Liberman, 1980).

In addition to these various perceptual interactions between acoustic or phonetic segments in stimuli resembling coherent speech, perceptual dependencies between successive isolated syllables have been demonstrated in a large number of studies. These sequential effects, too, occur both forward and backward in time. To quote two recent examples: Repp, Healy, and Crowder (1979) have shown that two isolated vowels presented in close succession influence each other's perception, with backward effects being at least as strong as forward effects; a similar result for pairs of CV syllables has been reported by Diehl, Elman, and McCusker (1978).

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR01996 to the Haskins Laboratories. I am grateful to Sarah Peck and Patti Jo Price for help in running subjects and analyzing the data, and to Virginia Mann for helpful comments.

Perceptual dependencies between isolated stimuli of the same class are typically contrastive in nature and have been attributed to response bias (Diehl, Lang, & Parker, 1980). The contextual effects occurring in single coherent speech stimuli, on the other hand, often involve interactions between segments from different classes (e.g., fricatives and vowels) and therefore cannot be so easily attributed to response bias (even though the effects are typically found to be contrastive if the segments involved have a dimension in common, such as place of articulation). Rather, they invite explanations in terms of perceptual compensation for coarticulatory dependencies between the segments in question (Mann & Repp, 1980, in press). The present studies are concerned with a situation that straddles the boundary between the two types just discussed, as it concerns successive syllables of a similar type that may or may not be considered part of a single utterance, depending on their temporal relationship.

The effects investigated here were first demonstrated by Repp (1978: Exps. V & VI): In disyllabic synthetic utterances of the type VC_1-C_2V--where C_1 and C_2 are voiced stop consonants (either /b/ or /d/) cued, respectively, by formant transitions in and out of a silent closure interval--the perception of C_1 depends on C_2 and vice versa, at least when the cues for one or both are ambiguous with respect to place of articulation. The nature and extent of the perceptual interaction between C_1 and C_2 (or their respective cues) vary with the duration of the silent closure interval between the two signal portions corresponding to VC_1 and C_2V. A schematic illustration of this dependency is provided in Figure 1, which is taken from Repp (1978) and based on rather preliminary data.

Consider first the solid function labeled B (for "backward"), which represents the effect of C_2 on C_1. At closure durations below approximately 70 msec, listeners generally do not perceive C_1, i.e., they do not interpret the formant transitions leading into the closure as cues for a separate phonetic segment, even when those transitions specify a different place of articulation than the transitions out of the closure (see also Abbs, 1971; Dorman, Raphael, & Liberman, 1979; Repp, 1979). One way of describing this effect is to say that C_2 exerts a strong assimilative effect on C_1--the cues for C_1 are interpreted in conformity with the cues for C_2 and integrated with the latter into a single phonetic percept. As closure duration is increased beyond 70 msec up to about 200 msec, C_1 emerges as a separate phonetic percept if the formant transitions into the closure can be interpreted as specifying a place of articulation different from that of C_2. (Otherwise, a single stop consonant is heard, at the place of articulation common to C_1 and C_2.) At these closure durations, C_2 exerts a contrastive effect on the perception of C_1, i.e., an ambiguous C_1 tends to be assigned to a category different from C_2. Figure 1 shows that this backward contrast effect declines as closure duration is extended beyond 200 msec. At these long closure durations, listeners tend to hear C_1 and C_2 as separate phonemes even if they have the same place of articulation; in this latter case, double (geminate) stop consonants are heard. Essentially, this implies that VC_1 and C_2V are perceived as separate utterances, and it is reasonable that such a percept should be accompanied by a reduction or even disappearance of contrast effects.

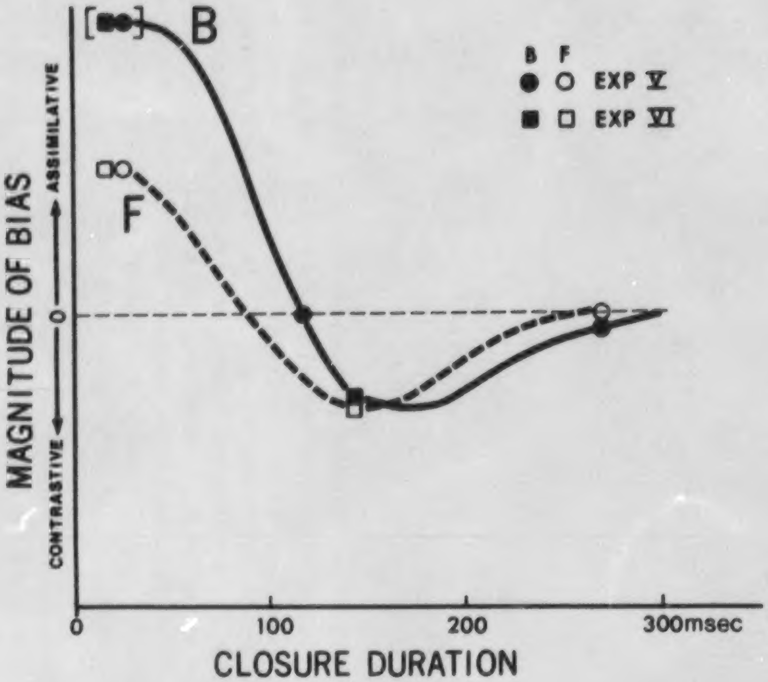


Figure 1. Schematic illustration of the perceptual interactions between C_1 and C_2 as a function of of closure duration. B = backward, F = forward. From Repp (1978).

Consider now the dashed function labeled F (for "forward") in Figure 1. It represents the influence of C_1 on the perception of C_2. The initial portion of this function is of special interest: As pointed out above, C_1 is not perceived as a separate phoneme at very short closure durations. However, Repp (1978) found some evidence that the formant transitions into the closure nevertheless had a perceptual effect--they biased responses toward the place of articulation they specified; thus, their effect on the perception of C_2 may be described as assimilative. In other words, their weight in the perceptual integration of the cues to C_1 and C_2 is not zero. At intermediate closure durations, however, where C_1 and C_2 are heard as separate phonetic segments (if perceived as different phonemes), C_1 exerts a contrastive effect on the perception of C_2. This forward contrast seems to be similar in magnitude to the backward contrast effect of C_2 on C_1; it, too, declines as closure duration is extended beyond 200 msec.

As can be seen from the few data points in Figure 1, Repp's (1978) experiments provided only a very rough sampling of the closure duration continuum. The schematic functions in the figure should be taken as hypotheses about the possible time course of assimilative and contrastive effects. It was the purpose of Experiment 1 to map out those functions in considerably more detail.

EXPERIMENT 1

All results represented in Figure 1 were obtained in blocked conditions, i.e., closure duration was held constant within a given test. This had the consequence that a simple bias to report two different consonants rather than only a single consonant could not be distinguished from true perceptual contrast. This problem was partially avoided in the present study by randomly varying closure duration within a certain range. If the perceptual dependency between C_1 and C_2 changes as a function of closure duration, this change cannot be attributed to response bias. If it does not change, on the other hand, it may be due to a response bias, as indeed a changing effect may be superimposed on such a bias. However, this was not considered a serious problem, in part because simple response bias was not expected to play an important role, and in part because systematic response bias--contrary to its bad reputation--is itself of theoretical interest.

For practical reasons, Experiment 1 was divided into three parts (1a, 1b, 1c), each covering one third of the total range of closure durations (10-310 msec). Experiment 1b was conducted some time before 1a and 1c.

Method

Subjects. Experiment 1b employed 12 subjects; they included nine paid student volunteers with varying experience in listening to synthetic speech, two research assistants, and the author. Experiments 1a and 1c employed nine subjects each, seven of whom participated in both experiments. Only two subjects (the author and one research assistant) participated in all three experiments.

Stimuli. The stimuli consisted of two synthetic stimulus continua, generated on the OVE IIIc synthesizer at Haskins Laboratories. The *VC* continuum consisted of seven stimuli ranging from /ab/ to /ad/ and differing only in the final formant transitions. The F_1 transition had a constant offset frequency of 541 Hz but changed in duration from 90 msec in stimulus 1 to 30 msec in stimulus 7. The F_2 and F_3 transitions had a constant duration of 50 msec but varied in offset frequency: F_2 offset changed from 1060 Hz in stimulus 1 to 1297 Hz in stimulus 7, and F_3 offset changed from 2181 Hz in stimulus 1 to 2539 Hz in stimulus 7, both in roughly equal steps. All transitions were stepwise-linear in 10-msec time segments. The formant frequencies of the initial steady-state portion were 777 Hz (F_1), 1147 Hz (F_2), and 2466 Hz (F_3). All VC stimuli had a duration of 180 msec, a constant fundamental frequency of 120 Hz, and an amplitude contour that increased over roughly two thirds of the stimulus and then declined.

The CV continuum consisted of seven stimuli ranging from /ba/ to /da/ and differing only in the initial transitions of F_2 and F_3. The F_1 transition was constant with an onset frequency of 459 Hz. F_2 onsets ranged from 1099 Hz in stimulus 1 to 1635 Hz in stimulus 7, and F_3 onset ranged from 2262 Hz in stimulus 1 to 2500 Hz in stimulus 7, both in roughly equal steps. All transitions were 50 msec long. The formant frequencies of the final steady-state portion were 728 Hz (F_1), 1156 Hz (F_2), and 2466 Hz (F_3). All CV stimuli had a duration of 290 msec, a fundamental frequency that was constant at 120 Hz over the first 90 msec and then fell linearly to 100 Hz, and an amplitude contour that rose slightly over the first 50 msec and then fell gradually until stimulus offset.

All stimuli were digitized at 10 kHz using the Haskins Laboratories pulse code modulation (PCM) system. Experimental sequences were recorded on magnetic tape using a special sequencing program. In each experiment, there were two conditions: a forward condition and a backward condition. In the forward condition, each of the seven stimuli from the CV continuum was preceded by one of the two endpoint stimuli of the VC continuum, at various interstimulus intervals that are referred to here as closure durations. In the backward condition, each of the seven stimuli from the VC continuum was followed by one of the two endpoint stimuli of the CV continuum, with various closure durations in between. Thus, there were 14 basic stimulus combinations in each condition. To obtain more observations for ambiguous stimuli, a 1-2-3-3-3-2-1 frequency distribution was imposed on the seven-member continua, so that the basic test unit contained 2 x (1 + 2 + 3 + 3 + 3 + 2 + 1) = 30 stimuli. In each experiment, each VC-CV stimulus occurred with five different closure durations, in a random sequence containing 5 x 30 = 150 stimuli. Three such sequences of 150 stimuli were recorded on each experimental tape. The interval between successive VC-CV combinations was 3 sec.

The three experiments differed only in the range of closure durations. Within each experiment, closure durations varied in 25-msec steps over a 100-msec range. Experiment 1a covered the range from 10-110 msec, Experiment 1b that from 110-210 msec, and Experiment 1c that from 210-310 msec.

In addition, randomized sequences of isolated VC and CV syllables were recorded. Each of these two sequences contained 75 stimuli, resulting from five replications of the basic 15-stimulus unit due to the 1-2-3-3-3-2-1 frequency distribution of the 7 stimuli on each continuum. The interstimulus interval was 2 sec. These tapes were used in all three experiments.

Procedure. Each experiment required two sessions per subject of approximately 90 minutes duration. At the beginning of each session, the subject listened to the isolated CV and VC sequences, in that order. Then the forward and backward tapes were presented. Their order was counterbalanced between subjects and reversed between the first and second sessions. In each experiment, the most ambiguous stimuli (i.e., stimuli 3-5 from a given continuum) received a total of 30 responses from each subject when presented as isolated monosyllables and 18 responses when presented in a specific VC-CV combination.

The response choices given to the subjects were the following: B and D for isolated syllables; B, D, BD, and DB for VC-CV combinations. In Experiment 1c, the choices B and D for VC-CV combinations were changed to BB and DD, respectively, since the closure durations were in the range where listeners were expected to hear geminate stops. The listeners were never required to distinguish between single (B, D) and geminate (BB, DD) stops; although such a distinction may have provided useful information, it was felt that it would have made the task too complicated. Although listeners were encouraged to note down any other consonants heard, there were hardly any occurrences of responses other than B and D and their combinations.

The tapes were played back at a comfortable intensity on an Ampex AG-500 tape recorder, and the subjects listened binaurally over TDH-39 earphones in a quiet room. The listeners were fully informed about the structure of the stimuli before each condition.

Results and Discussion

A gross measure of the perceptual interaction between C_1 (VC) and C_2 (CV) is provided by

$$\begin{aligned}&[(100/n)\textstyle\sum_i(\text{responses of D or DD, DB) to }VC_i\text{-/ba/}]\\ -\ &[(100/n)\textstyle\sum_i(\text{responses of D or DD, DB) to }VC_i\text{-/da/}]\end{aligned}$$

in the backward condition, and by

$$\begin{aligned}&[(100/n)\textstyle\sum_i(\text{responses of D or DD, BD) to /ab/-}CV_i]\\ -\ &[(100/n)\textstyle\sum_i(\text{responses of D or DD, BD) to /ad/-}CV_i]\end{aligned}$$

in the forward condition, where i indexes the seven stimuli on a given synthetic continuum and n is the total number of responses to the stimuli on a given continuum. Thus, the index is a percentage difference and varies from -100 for maximal contrast to +100 for maximal assimilation. These indices of stimulus interaction are plotted as a function of closure duration in Figure 2, separately for the forward and backward conditions.

In Experiment 1a (Fig. 2a), there was a strong assimilative backward effect at the shortest closure durations, as expected. It reflects the strong tendency to perceive only a single stop consonant that corresponds to C_2. As the closure duration increased, the backward effect changed rapidly from assimilative to contrastive, with the crossover occurring at about 55 msec of closure duration. Although such a crossover had been predicted, it occurred considerably earlier (i.e., at a shorter closure duration) than expected on

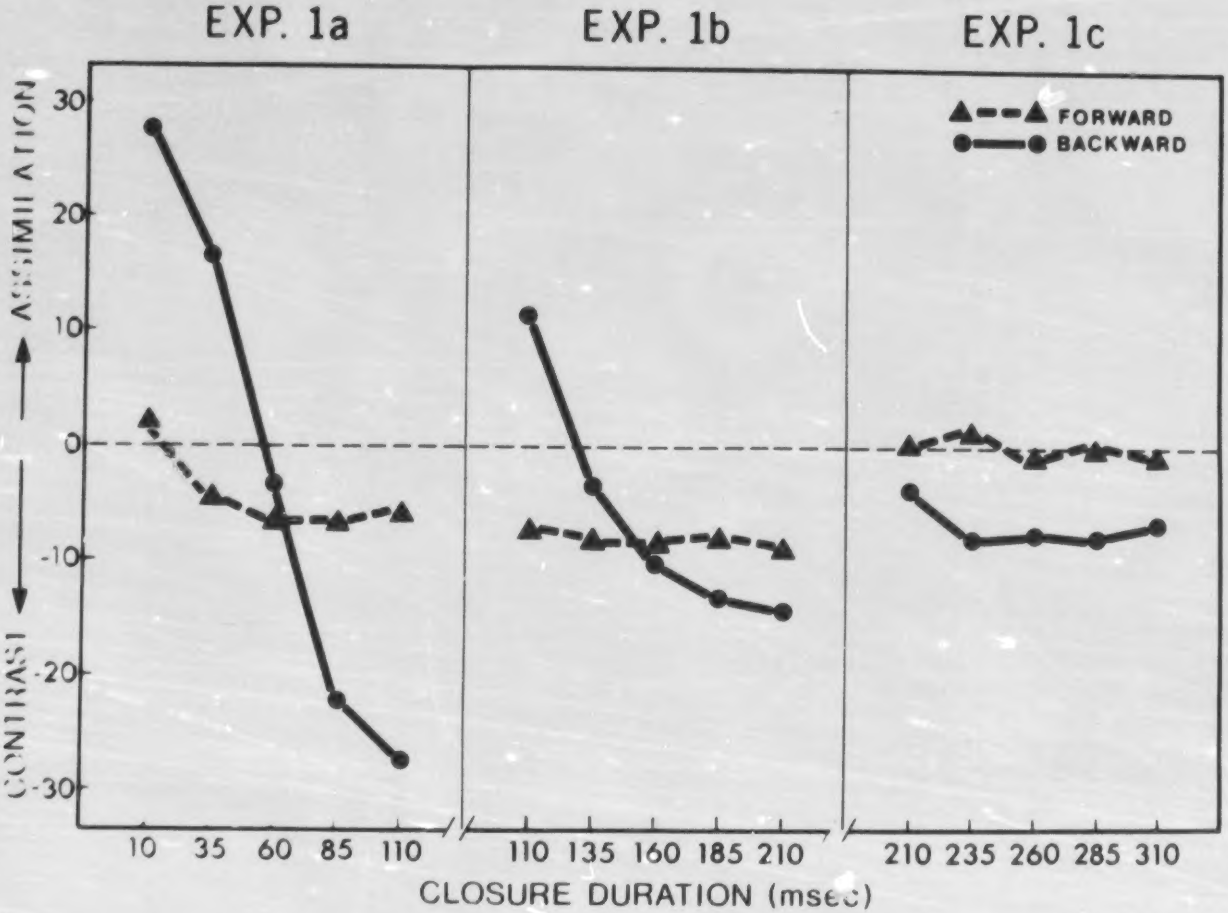


Figure 2. Forward and backward interactions between C_1 and C_2 as a function of closure duration (Exp. 1).

the basis of earlier data (cf. Figure 1). The crossover marks the emergence of C_1 as a separate phonetic percept (if different from C_2), and the contrastive effect indicates that there was a strong tendency to perceive C_1 as different from C_2.

The forward function in Experiment 1a, on the other hand, was considerably flatter than the backward function. In an analysis of variance, this was reflected in a highly significant interaction between the effects of Condition (forward vs. backward) and Closure Duration, $F(4,32) = 21.1$, $p < .001$, in addition to a highly significant main effect of Closure Duration, $F(4,32) = 27.1$, $p < .001$, which was primarily due to the backward function. There was a constant small forward contrast effect at closure durations beyond 35 msec; only at the shortest closure duration (10 msec), there was a minuscule assimilation effect. The change in the forward effect with closure duration was significant in a separate test, $F(4,32) = 5.4$, $p < .01$. However, the assimilative effect at the shortest closure duration was not significantly different from zero; it was shown by only five out of nine subjects. Repp (1978) found that the cues for C_1 influenced perception even though C_1 was not perceived as a separate phoneme. The present results provide only weak support for this earlier observation, as there was no absolute assimilative forward effect, only a relative reduction in the contrast evident at longer closure durations.

Let us turn now to Experiment 1b (Fig. 2b), which examined the region of intermediate closure durations. The backward function can be seen to follow very much the predicted course (cf. Figure 1): An assimilative effect at the shortest closure duration (110 msec) shifted toward a pronounced contrastive effect at longer closure durations, with the crossover occurring at about 130 msec of closure duration. No return to the zero baseline was indicated at the longest closure duration, suggesting a temporal range of the backward effect substantially exceeding 210 msec--an unexpected finding. In contrast to the backward function, the forward function was completely flat, showing a moderate contrast effect at all closure durations. The different shapes of the functions were reflected in a highly significant interaction of the effects of Condition and Closure Duration, $F(4,44) = 16.2$, $p < .001$, in addition to a significant main effect of Closure Duration, $F(4,44) = 20.6$, $p < .001$, which was solely due to the backward function. There was no significant effect of Closure Duration on the forward effect, as determined in a separate test, $F(4,4) = 0.5$.

The most unexpected result was the large discrepancy between the backward effects for the same closure duration (110 msec) in Experiments 1a and 1b: In Experiment 1b there was an assimilative effect, whereas, in Experiment 1a, there was a contrast effect that actually exceeded the contrast effect at the longest interval (210 msec) in Experiment 1b. Instead of a single crossover from positive to negative backward effects (expected to be at approximately 115 msec, according to Figure 1), there were two: one at 55 msec in Experiment 1a, and the other at 130 msec in Experiment 1b. These results are indicative of strong stimulus range effects (due to the range of closure durations used in a given condition) on the listeners' perception of the stimuli--more precisely, on their tendency to hear one vs. two (different) stop consonants (cf. Repp, 1980a). Indeed, single-consonant responses to conflicting sets of C_1 and C_2 cues did not occur at the 110-msec interval in

Experiment 1a, but appeared with some frequency at the same interval in Experiment 1b.

In Experiment 1c (Fig. 2c), the backward effect was contrastive throughout, but there was a significant reduction in contrast at the shortest interval (210 msec), $F(4,32) = 4.7$, $p < .01$, reminiscent of the more pronounced trends in the backward functions of Experiments 1a and 1b. The forward condition, on the other hand, showed neither any contrast nor any effect of closure duration. The difference between forward and backward effects was significant, $F(1,8) = 8.3$, $p < .05$. The different magnitudes of the backward contrast effects at 210 msec in Experiments 1b and 1c again suggest a stimulus range effect. The cause of the difference in the amount of forward contrast between the two experiments is less clear; perhaps, the difference in response choices (B and D vs. BB and DD) played a role.

Despite the unexpectedly strong stimulus range effects, the additional influence of closure duration is clearly evident in Figure 1. Backward contrast at the respectively longest intervals in each range (110, 210, 310 msec) declined as closure duration increased, suggesting that the effect might disappear when closure durations reach 400-500 msec. A "neutral" estimate of the closure duration where backward contrast emerges might be 100 msec; the corresponding point for forward contrast might be 20 msec. Forward contrast seemed to disappear earlier and was definitely less pronounced than backward contrast. On the whole, these results confirm Repp's (1978) earlier observations; however, backward contrast and stimulus range effects were considerably stronger than expected, and no forward assimilation effect was obtained at short closure durations.

A more detailed examination of the frequencies of the various responses to the individual stimulus combinations and to the isolated VC and CV syllables is presented in the Appendix.

EXPERIMENT 2

It was pointed out in the introduction to Experiment 1 that any effect (assimilative or contrastive) that remained constant within an experiment, such as the forward contrast in Experiment 1b, may have been due to response bias. Such a bias may have been contingent on the identification of C_2: The listeners may have first categorized C_2 and then followed their biases in deciding whether to respond C_2 or C_1C_2. Underlying such a bias may have been the motivation to identify as many consonants as possible, even though the subjects were instructed to write down just what they heard.

There were reasons to believe that many, if not all, of the effects demonstrated in Experiment 1 were perceptual in origin: the changes with closure duration within and across the three sub-experiments, the effects of acoustic stimulus structure (see Appendix), the generally high consistency among subjects, and the fact that the author--who presumably followed the instructions without any bias--showed most of the effects described. Still, the extent of the influence of stimulus range was alarming, as it suggests a change in response criteria. Clearly, the perceptual distinction between single stops and a sequence of two stops is not very stable (cf. also Repp, 1980a) and, therefore, must be highly susceptible to response bias. For this reason, Experiment 2 was conducted to see whether forward and backward contrast effects would be obtained in a discrimination task, where response bias presumably plays little or no role.

Because of practical limitations, only two closure intervals could be selected (150 and 250 msec), both in the region where contrast effects were expected. The task was set up so that listeners had to distinguish between members of the VC or CV continuum, in isolation and in the presence of one or the other post- or precursor (the endpoints of the other continuum). It is well known that, on such synthetic stimulus continua, discrimination performance is high when the two stimuli to be compared fall on opposite sides of the category boundary, but very low when the two stimuli are from the same phonetic category. This is the familiar pattern of categorical perception. In the present study, the question was whether a pre- or postcursor would shift the discrimination peak and/or change within-category discrimination performance on a given continuum. If the effect is contrastive, as expected, the peak should shift *towards* the category represented by the pre- or postcursor and/or discrimination performance should be improved within that category.

Method

Subjects. Sixteen subjects participated, including fourteen paid volunteers, one research assistant, and the author.

Stimuli and design. The stimuli were the same as in Experiment 1. There were 12 experimental conditions, resulting from the orthogonal combination of three factors: backward vs. forward (i.e., VC vs. CV discrimination), closure duration (150 vs. 250 msec), and context (none vs. /b/ vs. /d/ pre- or postcursor). To facilitate the discrimination task, none of the factors was randomized. As in Experiment 1, the pre- or postcursors were the endpoint stimuli from the VC and CV continuum, respectively. Thus, in the forward condition, the subjects' task was to discriminate stimuli from the CV continuum in isolation and when preceded by either /ab/ or /ad/ at a given closure duration; in the backward condition, they had to discriminate stimuli from the VC continuum in isolation and when followed by either /ba/ or /da/.

The stimuli to be discriminated were arranged in AXB triads, with interstimulus intervals of 500 msec in the pre- or postcursor conditions. Isolated VC or CV stimuli were separated by as much silence as equaled their temporal separation in the corresponding pre- or postcursor conditions (950 or 1050 msec for VC stimuli and 840 or 940 msec for CV stimuli, depending on the closure duration condition). The interval between AXB triads was 3 sec in all cases.

The stimulus differences to be detected were two-step separations on the seven-member synthetic continua. Thus, there were five different contrasts (1-3, 2-4, 3-5, 4-6, 5-7) each of which appeared in four possible AXB arrangements (AAB, ABB, BAA, BBA), resulting in twenty triads which were repeated five times in random order to give a total of 100. Each of the twelve experimental conditions contained such a set of 100 triads, preceded by four easy practice triads that served to illustrate the structure of the stimuli.

Procedure. Each subject participated in four one-hour sessions. The four conditions resulting from the orthogonal combination of the forward-backward and closure duration factors were presented on separate days in an order that was counterbalanced across subjects according to a Latin-square schedule. In each session, the isolated VC or CV condition was presented first; it served both as familiarization and as a baseline for comparison with the pre- or postcursor conditions that followed. The order of the following /b/ and /d/ pre- or postcursor conditions was counterbalanced across subjects.

The equipment was the same as in Experiment 1. The subjects indicated their choices by writing A or B, depending on whether the second stimulus sounded more similar to the first or to the third, guessing if necessary. All subjects were fully informed about the structure of the stimuli and knew where the difference was located.

Results and Discussion

The results are shown in Figure 3, the forward condition (CV discrimination) at the top and the backward condition (VC discrimination) at the bottom. The discrimination functions for isolated stimuli (dotted, triangles) had the familiar peaked shape. These results served only as a guideline and were not included in the statistical analysis. Performance in the pre- and postcursor conditions was slightly lower than for isolated stimuli, indicating a small amount of interference due to the added stimulus component.

The main results are easy to summarize. In no case was there a shift in the discrimination peak as a function of pre- or postcursor condition. However, discrimination performance tended to be improved at the end of the continuum that corresponded to the category represented by the pre- or postcursor--a pattern indicative of a contrast effect. This effect, revealed as an interaction between the (highly significant) effect of position on the continuum and the effect of /b/ vs. /d/ pre- or postcursor, was significant both in the forward condition, $F(4,60) = 6.2$, $p < .001$, and in the backward condition, $F(4,60) = 2.6$, $p < .05$. Neither effect was influenced by closure duration.

These results confirm the existence of perceptual contrast effects between C_1 and C_2, in both directions. The effects were perhaps smaller than those observed in the identification task (Experiment 1), since they were not sufficient to shift discrimination peaks. However, whereas the contrast effects in Experiment 1 may have been augmented by response bias, the present contrast effects definitely cannot be ascribed to such a bias. The present results differ from those of Experiment 1 in that forward contrast was larger and more reliable than backward contrast, and in that neither effect decreased as the closure duration was extended from 150 to 250 msec. These discrepancies cannot be explained at present.

GENERAL DISCUSSION

As was pointed out in the Introduction, there are two candidate explanations for the contrast effect reported here: (1) These effects may be related to the sequential effects observed in studies of selective adaptation and anchoring, and thus may represent either an auditory interaction or a response

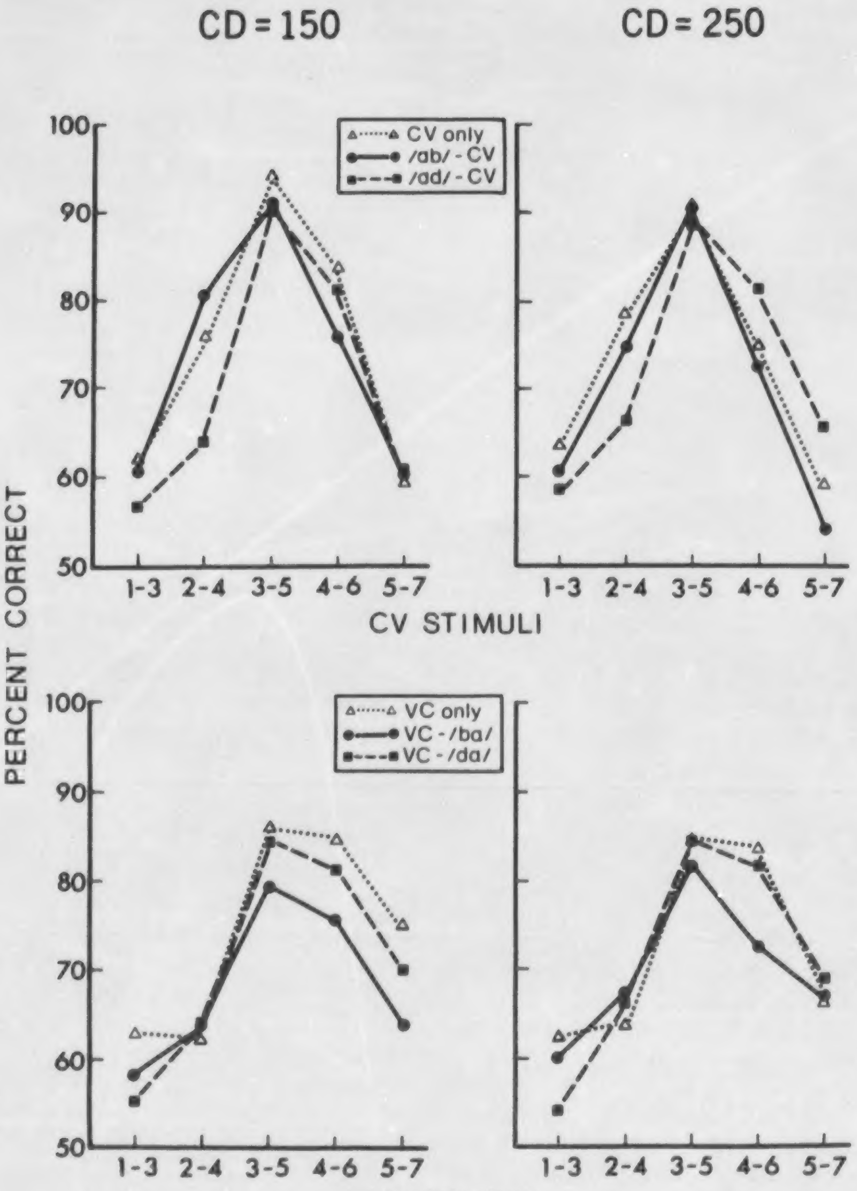


Figure 3. AXB discrimination performance for VC and CV stimuli in isolation and in context, as a function of context stimulus and closure duration (Exp. 2).

contrast phenomenon. (2) The effects may reflect a perceptual compensation for assimilatory coarticulatory effects in the production of sequences of two stop consonants.

Let us consider the first class of hypotheses. The results of Experiment 2 seem to rule out response contrast as a valid explanation, although such a mechanism may have played a supplementary role in Experiment 1. This leaves us with some form of auditory interaction as the possible cause of the contrast effects. It is relevant here to consider some results from studies of selective adaptation. Even though adaptation studies present precursor stimuli many times rather than just once, the close temporal contiguity of VC and CV components in the present studies may have produced some adaptation (i.e., auditory contrast). However, Ades (1974) found no cross-adaptation between VC and CV syllables. Later, Pisoni and Tash (1975) and Sawusch (1977) showed that the syllable-final formant transitions of VC-like stimuli can have an adaptation effect on CV stimuli; however, the direction of this effect reflects auditory similarity, not phonetic similarity. Since /ab/ and /ba/ are approximate mirror images (hence, not similar) in auditory terms, the auditory adaptation effect corresponds to an assimilation effect in phonetic terms and thus runs counter to the contrast effects found in the present studies. Sawusch (1977) suggested that the reason why /ab/ does not adapt /ba/ may be that an auditory adaptation effect is canceled by a simultaneous phonetic adaptation effect in the opposite direction. However, since "phonetic adaptation" is essentially the same as response contrast, this hypothesis cannot fully explain the present results.

Any explanation in auditory terms must deal with the findings that backward contrast is at least as strong as forward contrast, that the contrast effects depend on the duration of the closure interval (at least in an identification task), and that stimulus range has a very large effect. While it is difficult to rule out auditory explanations altogether at this stage, it is not clear how such an explanation could account for all aspects of the present findings.

Consider now the alternative hypothesis, that speech perception reflects speech production. According to one rather specific version of this hypothesis, perceptual contrast compensates for coarticulation. At the time of Repp's (1978) studies, such an explanation was not considered because the place of articulation of stop consonants such as /b/ and /d/ was not thought to be subject to coarticulatory shifts. In the meantime, however, we have obtained clear evidence of such shifts for stops following fricatives (Repp & Mann, in press) and liquids (Mann, in press). Thus, it seems not only conceivable but even likely that a preceding stop would influence the articulation of a following stop. Similarly, a following stop might affect the articulation of a preceding stop. In other words, there may be bidirectional coarticulation in sequences of two stop consonants, and since coarticulation is by definition assimilative in nature, perceptual compensation for such an effect would lead to contrast effects. Further perceptual studies using natural speech, as well as acoustic analyses of natural utterances, are now in progress to confirm the existence of coarticulatory shifts in place of articulation of two stops produced in sequence.

Reference to speech production simplifies considerably the interpretation of the present results. It explains not only the existence of contrast effects at closure durations longer than about 100 msec but also the existence of assimilation effects at shorter closure durations. Rather than reflecting some general principle of auditory processing, the change from contrast to assimilation as closure duration is shortened is likely to be related to the fact that closure durations become too short for the articulation of two stops in sequence (cf. Dorman et al., 1979). A typical average closure duration for two-stop sequences in isolated disyllables is about 180 msec (Westbury, Note 2; Repp, 1980b), whereas the typical closure duration for single intervocalic stops is about 80 msec (Westbury, Note 2; Umeda, 1977). Thus, listeners tend to hear only a single stop consonant at short closure durations because the closure duration acts as a cue to the class "single stop".

This argument works also in the other direction: Longer closure durations cue the class "two stops", and therefore listeners tend to report two different stops. Indeed, this hypothesis is sufficient to explain why contrast effects occur: If the closure duration is long enough to indicate a two-stop sequence, listeners will naturally try to interpret the place-of-articulation cues in the VC and CV portions in different ways. Thus, assimilation and contrast effects can be explained on an articulatory basis, whether or not two-stop sequences actually exhibit coarticulatory shifts in production. However, the demonstration of such shifts would place an articulatory interpretation on even firmer ground.

Even the stimulus range effects observed in Experiment 1 can be explained by reference to articulation. To determine whether a given closure duration is short or long, listeners presumably take the prevailing rate of articulation into account. If the range of closure durations includes only relatively short intervals, then the utterances will seem to be spoken at a fast rate, and a shorter interval will be required to separate one-stop from two-stop percepts than when the range of closure durations includes only relatively long intervals. Thus, range effects can be interpreted as a perceptual adaptation to changes in perceived speaking rate.

In summary, it seems that reference to speech production provides an explanatory framework that is more elegant, parsimonious, and ecologically valid than hypotheses framed exclusively in terms of general auditory mechanisms. While auditory processes certainly play a role in the initial stages of processing--and indeed may account for some aspects of the present data--the conclusion that speech perception is guided by principles of speech production and by listeners' internal representations of the resulting characteristic acoustic patterns seems inescapable in the light of accumulating evidence.

REFERENCE NOTES

1. Kunisaki, O., & Fujisaki, H. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics*, 1977, *11*, 85-91.
2. Westbury, J. R. *Temporal control of medial stop consonant clusters in English*. Paper presented at the 93rd Meeting of the Acoustical Society of America in Austin, Texas, April 1977.

REFERENCES

Abbs, M. H. *A study of cues for the identification of voiced stop consonants in intervocalic contexts*. Unpublished doctoral dissertation, University of Wisconsin, 1971.

Ades, A. E. How phonetic is selective adaption? Experiments on syllable position and vowel environment. *Perception & Psychophysics*, 1974, *16*, 61-67.

Diehl, R. L., Elman, J. L., & McCusker, S. B. Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 599-609.

Diehl, R. L., Lang, M., & Parker, E. M. A further parallel between selective adaptation and contrast. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 24-44.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 1979, *65*, 1518-1532.

Mann, V. A. Effect of preceding liquid on stop consonant perception. *Perception & Psychophysics*, in press.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 1980, *28*, 213-228.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, in press.

Pisoni, D. B., & Tash, J. B. Auditory property detectors and processing place features in stop consonants. *Perception & Psychophysics*, 1975, *18*, 401-408.

Raphael, L. J., Dorman, M. F., & Liberman, A. M. On defining the vowel duration that cues voicing in final position. *Language and Speech*, 1980, *23*, 297-307.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, 1978, *24*, 471-485.

Repp, B. H. Influence of vocalic environment on perception of silence in speech. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-57*, 267-290.

Repp, B. H. A range-frequency effect on perception of silence in speech. *Haskins Laboratories Status Report on Speech Research*, 1980, *SR-61*, 151-166. (a)

Repp, B. H. Perception and production of two-stop-consonant sequences. *Haskins Laboratories Status Report on Speech Research*, 1980, *SR-63/64*, this volume. (b)

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 129-145.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 621-637.

Repp, B. H., & Mann, V. A. Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, in press.

Sawusch, J. R. Processing of place information in stop consonants. *Perception & Psychophysics*, 1977, *22*, 417-426.

Umeda, N. Consonant duration in American English. Journal of the Acoustical Society of America, 1977, 61, 846-858.

APPENDIX

Figure 4 shows selected data from the backward conditions in the three parts of Experiment 1. Each panel plots the percentage of D responses to VC syllables in isolation, and the combined percentages of D (DD) and DB responses separately for VC-/ba/ and VC-/da/ stimuli, each as a function of stimulus changes along the VC continuum. (Effectively, the figure shows DB responses to VC-/ba/ and D (DD) responses to VC-/da/, since D (DD) responses to VC-/ba/ and DB responses to VC-/da/ were extremely rare, as were other "irregular" responses.) The VC stimuli in isolation exhibited a rather sharp category boundary between stimuli 4 and 5, as can be seen in all panels of the figure.

Figure 4a shows the results at the shortest closure duration (10 msec) of Experiment 1a. If C_2 had completely dominated C_1 at this brief interval, the response functions should have been completely flat: 100 percent B (i.e., 0 percent DB) responses to all VC-/ba/ stimuli, and 100 percent D responses to all VC-/da/ stimuli. Clearly, this was not the case. Even at this short closure duration, there was a substantial percentage of two-consonant responses, DB in the case of VC-/ba/ stimuli and BD in the case of VC-/da/ stimuli. BD responses, which are represented in Figure 4a by the difference of the VC-/da/ function from 100 percent, were more frequent than DB responses (reaching 50 percent vs. only 33 percent), indicating that the /b/ in /ab/ (VC stimuli 1-3) followed by /da/ was easier to "detect" than the /d/ in /ad/ (VC stimuli 5-7) followed by /ba/. This contradicts an earlier finding by Repp (1978), suggesting stimulus-specific differences. Note also that the "detectability" of C_1 cues was affected by the acoustic composition of the formant transitions: Two-stop responses were most frequent for the endpoint stimuli and decreased for stimuli close to the boundary.

Figure 4b shows a "close-up" of the strong contrast effect at a closure duration of 110 msec (Exp. 1a). One feature to note here is that the contrast effect was sufficiently strong to affect the endpoint stimuli of the VC continuum: /ab/ (VC stimulus 1) followed by /ba/ received 37 percent DB responses, and /ad/ (VC stimulus 7) followed by /da/ received 26 percent BD (74 percent D) responses. This may suggest a simple response bias in favor of two-consonant responses, but note that the frequency of these responses was strongly affected by acoustic changes in the VC stimulus: DB responses increased from 37 percent (VC stimulus 1) to 83 percent (VC stimulus 4), even though VC stimuli 1-4 were all identified as /ab/ in isolation, and BD responses increased from 26 percent (VC stimulus 7) to 62 percent (VC stimulus 5), even though VC stimuli 5-7 were all identified as /ad/ in isolation. This evidence argues strongly against a simple response bias as the only factor (although such a component may have been present) and instead implies that the listeners were sensitive to the precise trajectories of the VC formant transitions.

Figure 4c shows the results for the 110-msec interval in Experiment 1b, backward condition. The assimilative effect (of C_2 on C_1) obtained here

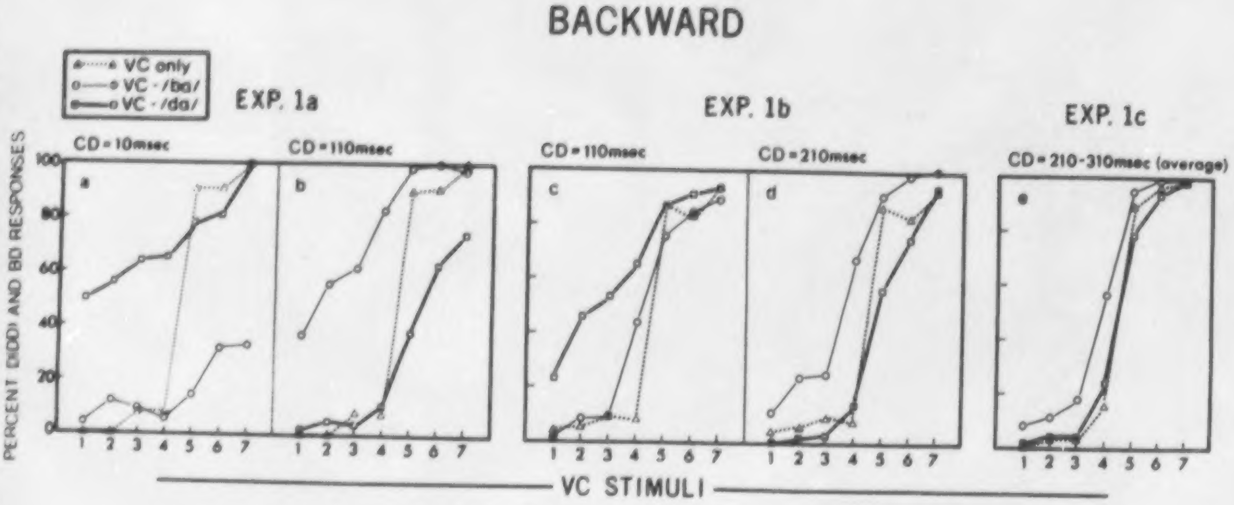


Figure 4. Response functions in various backward conditions (Exp. 1).

looked quite different from that shown in Figure 4a. Not only were the response functions steeper but the effect seemed to be almost entirely due to the /da/ postcursor. In other words, the cues in /da/ tended to dominate those in /ab/ (leading to many D responses), but /ba/ had little effect on /ad/ (leading to DB responses). A different way of looking at this asymmetry is to assume that VC syllables were generally perceived as more /ad/-like when followed by any CV syllable. This interpretation is preferred because the asymmetry continued at longer closure durations (shown in Figures 5d and 5e), where the perceptual interaction between C_1 and C_2 was contrastive. There, only the /ba/ postcursor seemed to exert an effect. Why listeners tended to hear more syllable-final Ds in VC-CV stimuli than in isolated VCs is not known, but it was apparently due to the specific stimuli used, since Repp (1978) found no such shift in his backward condition.

Figure 5 shows detailed results for the forward conditions. The plots are analogous to those in Figure 4, with the roles of VC and CV reversed. In Figure 5a, the results for the shortest closure duration (10 msec) are displayed. The dominance of C_2 over C_1 is reflected here by the relative steepness of the response functions for VC-CV combinations. The figure shows a tiny assimilative effect at the lower (/ba/) end of the CV continuum. Also, there was an asymmetry: D responses were more frequent with either VC precursor than with isolated CV syllables. Curiously, this asymmetry was reversed at longer closure durations (Figures 5b-5d), with listeners giving fewer syllable-initial D responses in VC-CV context than to isolated CVs. No such asymmetries had been found by Repp (1978).

Figure 5a does not show the percentages of two-consonant responses: At the 10-msec closure duration in the forward condition, there were 48 percent BD responses to /ab/ followed by /da/ (CV stimulus 7) and 31 percent DB responses to /ad/ followed by /ba/ (CV stimulus 1). These frequencies agree very well with the corresponding percentages (50 and 33 percent, respectively) for the identical endpoint stimulus combinations in the backward condition (cf. Figure 4a). It can now be understood why there was no significant assimilative forward effect at the shortest closure duration. Given the unexpectedly high rate of two-consonant responses, and given that such responses imply a contrast effect, whatever assimilative effect may have existed was cancelled by simultaneous contrast. For reasons that are not entirely clear, the stimuli with the 10-msec interval were perceived like earlier stimuli with a closure duration of 60 msec or so (cf. Repp, 1978; Dorman et al., 1979; see also Figure 1).

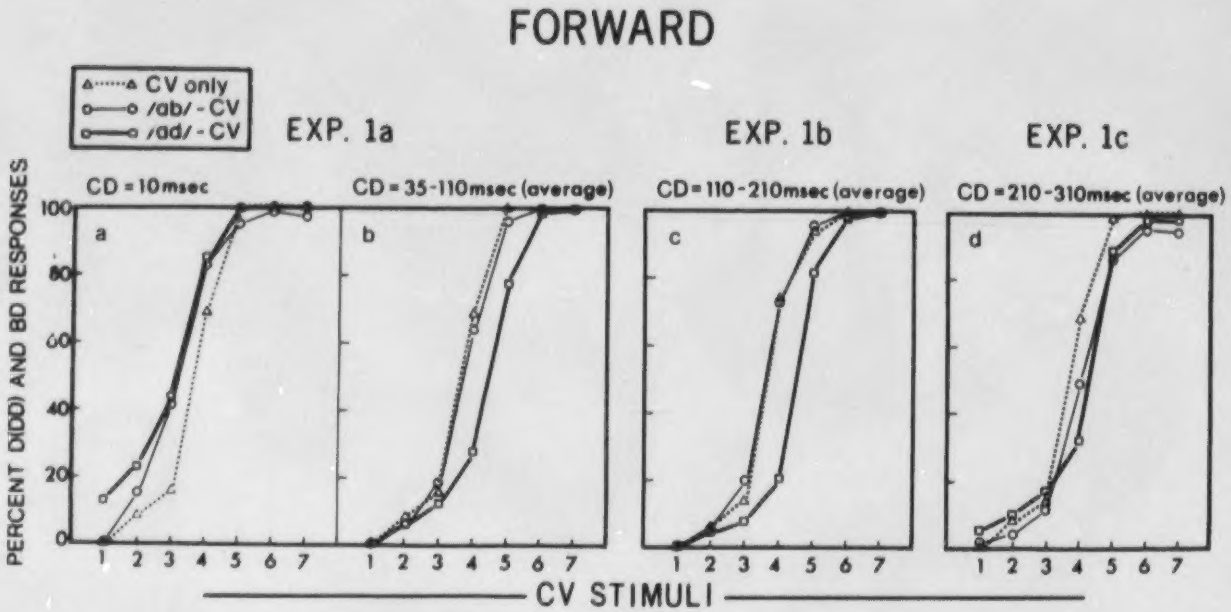


Figure 5. Response functions in various forward conditions (Exp. 1).

PERCEPTION AND PRODUCTION OF TWO-STOP-CONSONANT SEQUENCES

Bruno H. Repp

Abstract. The duration of the silent closure interval required to perceive two stop consonants in a VC_1C_2V sequence depends, to some extent, on their places of articulation. In production, too, the duration of the closure interval varies systematically with place. However, there appears to be little relation between the patterns of variability in production and in perception. Moreover, two analogous perceptual experiments--one using synthetic stimuli, the other, natural speech--yield quite different results. Thus, variations in the amount of closure required to perceive two successive stops seem to be governed by stimulus-specific acoustic factors, not by an internal representation of articulatory patterns or constraints. This conclusion is further supported by the unexpected finding that some listeners do not require any closure interval for accurate perception of both stops.

INTRODUCTION

Lisker (1957) first reported that, when the waveforms of naturally produced /ræg/ (with /g/ unreleased) and /bɪd/ are abutted without any intervening silence (which serves to indicate oral closure), listeners hear /ræbɪd/--that is, they fail to perceive the first (syllable-final) stop consonant. This effect was later rediscovered by Abbs (1971) and has, more recently, been investigated in considerable detail (Dorman, Raphael, & Liberman, 1979; Raphael & Dorman, in press; Repp, 1978, 1979a, 1979b, 1980; Rudnicky & Cole, 1978). These studies used both synthetic and natural speech, and a variety of stop-consonant combinations and vocalic contexts. Several studies assessed precisely what closure duration is needed between the VC_1 and C_2V waveforms to perceive both stop consonants on 50 percent of the trials; a typical value for this perceptual boundary on a continuum of varying silent closure durations is 70 msec. However, the explanation of the phenomenon is still far from clear.

Two basic possibilities may be distinguished. One is that the effect in question is entirely auditory; e.g., it might be due to interference of the cues for the second stop (the formant transitions out of the closure) with the processing of the cues for the first stop (the formant transitions into the

---

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories. Most of the work was done by my able assistant, Patti Price. I thank Gary Kuhn for permission to draw on his collection of VCV stimuli. Leigh Lisker, Virginia Mann, and Ignatius Mattingly contributed helpful comments.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

closure)--cf. Massaro (1975). If so, any variations across different stimuli in the amount of closure necessary for accurate perception of both stops should be explainable by reference to what is known about relevant auditory processes such as backward masking or gap detection. The other possibility is that perception mirrors articulation more or less directly, as appears to be the case with many other phenomena in speech perception. If so, then variations in the closure duration needed to perceive two stop consonants should be correlated with similar variations in the average (or, perhaps, the minimum) closure duration in naturally produced VC_1C_2V sequences. Neither of these alternatives has been unequivocally supported or rejected in recent studies of the influence of three primary auditory stimulus parameters (spectrum, duration, and amplitude of the two signal portions) on the location of the perceptual boundary (Repp, 1979a, 1979b). In part, this is due to an absence of systematic acoustic data based on natural productions, and to the consequent uncertainty as to the predictions of the "articulatory hypothesis".

The present paper remedies this situation by directly comparing perception and production of a set of utterances selected to be particularly relevant to the articulatory hypothesis. The set consists of the six possible sequences of the three voiced stop consonants of English, in vocalic context: /VbdV/, /VbgV/, /VdgV/, /VdbV/, /VgbV/, and /VgdV/. A preliminary study comparing perceptual boundary values (the closure duration needed to hear both stops, rather than only the second) for these six stimulus types was reported briefly by Liberman (1975). The stimuli in that experiment were synthetic and of the form /bɑC_1C_2ɑ/; the silent closure interval was varied from 0 to 125 msec in a number of steps. The results were quite clear: On one hand, stimuli in which place of stop articulation moved from front to back (/bd/, /bg/, /dg/) had boundary values of 75-90 msec; on the other hand, stimuli in which place of stop articulation moved from back to front (/db/, /gb/, /gd/) had boundaries between 0 and 25 msec of silence. These data pointed towards a possible articulatory basis: perhaps, back-front sequences are easier to articulate (and, hence, have shorter closures) than front-back sequences. However, no articulatory or acoustic observations were available that spoke to this suggestion.

Recently, Raphael and Dorman (in press) replicated the Liberman study using natural speech. In view of the fact that they used single tokens produced by a single speaker (stimuli nearly as unrepresentative as the synthetic tokens used by Liberman), the agreement with the results of the earlier study was striking. Front-back sequences again required 75-90 msec of closure for both stops to be heard; back-front sequences, on the other hand, had perceptual boundaries between 0 and 50 msec. Curiously, Raphael and Dorman did not raise the possibility of an articulatory basis for their results; instead, they briefly considered two psychoacoustic hypotheses, neither of which was well supported by their data. However, they acknowledged--as did Liberman (1975)--the need to replicate this pattern of results in vocalic contexts other than /ɑ-ɑ/.

This is one purpose of the present studies. It seems likely that any articulatory constraint relating to front-back vs. back-front movement in place of stop articulation would be essentially constant across different vocalic environments; therefore, if perception follows production--as the articulatory hypothesis asserts--the pattern of perceptual results, too, should be invariant across different vocalic contexts. In the less likely case that the articulatory dynamics of two-stop sequences strongly depend on the vocalic environment, the question becomes whether changing articulatory patterns correspond in any way to changing perceptual requirements as a function of vocalic context. If psychoacoustic factors are at work in the perceptual suppression of the first stop, considerable variability in the pattern of results might be expected across different vocalic contexts because the acoustic properties of the stimuli change radically with changes in the surrounding vowels; in particular, the formant transitions conveying the places of articulation of the two stops may change in extent, shape, and direction. According to the auditory hypothesis, however, the pattern of variability observed in perception should have little relation to what occurs in speech production.

Thus, the present studies address three issues: (1) Does the perceptual boundary indeed vary across different combinations of stops, as earlier studies suggest, and if so, is this pattern of results stable across different vocalic contexts? (2) Do closure durations in corresponding natural utterances vary across different combinations of stops, and if so, is this pattern stable across different vocalic contexts? (3) Is there any consistent relationship between the patterns observed in perception and in production?

EXPERIMENT 1: PERCEPTION--SYNTHETIC STIMULI

Method

Subjects. Eleven subjects participated. They included nine paid volunteers (mostly Yale undergraduates), one research assistant, and the author. All were native speakers of American English except for the author whose native language is German. Earlier studies indicated no systematic differences between his perception of VC_1C_2V stimuli and that of native speakers of English.

Stimuli. Because convincing unreleased syllable-final stops at all three places of articulation are difficult to synthesize following vowels other than /ɑ/, the vowel in the first syllable was always /ɑ/, and only the vowel in the second syllable was varied. The basic stimulus components were three VC syllables--/ɑb/, /ɑd/, and /ɑg/--and nine CV syllables: /bɑ/, /dɑ/, /gɑ/, /bi/, /di/, /gi/, /bu/, /du/, /gu/. All syllables were produced by the OVE IIIc serial resonance synthesizer at Haskins Laboratories. Out of convenience, the parameters were taken from a set of VCV utterances previously synthesized by a colleague using a computer procedure (CONVERT) which permits the conversion of parameters of natural-speech spectrograms into synthesizer parameter values. Thus, the synthetic syllables were simplified recreations of natural speech; the fact that they were derived from VCV (rather than VC_1C_2V) utterances seemed unimportant, especially since there were no obvious coarticulatory effects across the closure period (cf. Öhman, 1966) in the original utterances. Only periodic excitation was used in the synthetic stimuli.

The stimuli were regularized with respect to duration and fundamental frequency. All VC syllables were 180 msec long and had a constant fundamental of 120 Hz. All CV syllables were 290 msec long and had a fundamental frequency contour that began at 120 Hz, remained steady for 40-140 msec (depending upon the individual stimulus, as copied from natural speech), and then fell steadily to a value between 94 and 105 Hz. All amplitudes and formant trajectories remained as traced from natural speech. This implied lower output amplitudes for /Cu/ than for /Cɑ/ and /ɑC/ syllables, with /Ci/ amplitudes in between. (Repp, 1979b, showed that stimulus amplitude plays only a minor role in the paradigm used here.)

All synthetic stimuli were digitized at 10 kHz using the Haskins Laboratories PCM system. Three test tapes were then created, identical except for the vowel of the CV syllables (/ɑ/, /i/, /u/), which varied across tapes. Each tape contained first a randomized sequence of the six component syllables (/ɑb/, /ɑd/, /ɑg/, /bV/, /dV/, /gV/) in which each stimulus occurred 10 times, with interstimulus intervals (ISIs) of 3 sec. The stimuli in the main portion of the test consisted of the six possible /ɑC_1C_2V/ disyllables ($C_1 \neq C_2$), with silent closure intervals varying in ten 10-msec steps from 15 to 115 msec. The resulting 66 disyllabic stimuli were recorded in five different randomizations, with ISIs of 3 sec.

Procedure. The subjects listened in a quiet room over TDH-39 earphones. The tapes were played back at a comfortable intensity on an Ampex AG-500 tape deck. Each subject participated in two sessions. In each session, all three tapes were presented in counterbalanced order. Thus, each subject gave a total of 10 responses to each individual VC-CV stimulus combination, 20 responses to each isolated CV syllable, and 60 responses to each isolated VC syllable (since the same VC syllables occurred on each tape). The task was to identify by forced choice (in writing) all stop consonants heard. In the monosyllabic series, the response choices were "b", "d", "g"; the subjects were told that the stops could occur in either initial or final position. In the VC-CV series, there were nine response choices: "b", "d", "g", "bd", "bg", "dg", "db", "gb", "gd". The subjects were informed about the structure of these stimuli--that they were made up from the monosyllabic components just heard, with varying intervals of silence between them. They were also told that, at short intervals of silence, the first (syllable-final) stop tends to disappear from perception. They were asked to write down only what they heard, not to guess a supposed consonant that was not actually perceived.

Results and Discussion

Two subjects (paid volunteers) unexpectedly failed to hear a sufficient number of single stops in VC-CV combinations--they generally heard two stops, usually the correct ones, even when little or no silence was present. Their data were excluded, so that the following results are based on nine subjects.

Monosyllables. The identifiability of the stops in the isolated VC and CV components was good to excellent, considering the fact that most of the subjects had little experience with synthetic speech. The majority of the confusions was due to a few individual listeners who more or less consistently misidentified an individual stimulus. The /Ci/ set generated more confusions than the /Cɑ/, /Cu/, and /ɑC/ sets; the respective percentages of correct responses were 80.4, 98.0, 97.6, and 95.7.

VC-CV combinations: Two-stop vs. one-stop responses. The responses to VC-CV combinations were first scored in terms of two-stop vs. one-stop responses, regardless of whether the responses were correct (i.e., the equivalent of C_1, C_2, or C_1C_2) or not. (Exclusion of errors would have distorted the data because of certain systematic misidentifications, which are discussed below.) All VC-CV combinations showed the expected increase in two-stop responses as the silent interval increased in duration. The boundary values (50-percent cross-over points) for all but two of the labeling functions fell between 55 and 80 msec. Two functions, however, stood out--those for /ɑgbɑ/ and /ɑdbɑ/; these stimuli required much less silence for both stops to be heard and they received a nonnegligible number of two-stop responses even at the shortest silence duration. Note that both stimuli contain back-to-front movements of place of articulation, in agreement with Raphael and Dorman (in press).

Figure 1 summarizes the data in terms of percentage single-stop responses, averaged across all silence durations--a measure that takes into account differences in the lower and upper asymptotes of the response functions. (However, a plot in terms of boundary values yields a very similar pattern.) It can be seen that the deviant results for /db/ and /gb/ in the /Cɑ/ set have no parallel in the /Ci/ and /Cu/ sets; clearly, they are specific to the /Cɑ/ stimuli (to /bɑ/ in particular). The hypothesis that front-back sequences (the first three stimuli on the abscissa in Figure 1) would have lower boundary values (i.e., more single-stop responses) than back-front sequences (the last three stimuli on the abscissa) is not supported in the /Ci/ and /Cu/ sets, and only partially supported in the /Cɑ/ set, since /ɑg-dɑ/ did not have a low boundary value.

The deviant results for /ɑdbɑ/ and /ɑgbɑ/ led to highly significant effects in an analysis of variance. However, after exclusion of all /db/ and /gb/ stimuli from the analysis, there was no significant effect of either consonant combinations or vocalic context; the interaction of these two factors was marginally significant, $F(6,48) = 3.0$, $p < .05$, but difficult to interpret.

VC-CV combinations in the /Cɑ/ set tended to have somewhat shorter boundaries than those in the /Ci/ and /Cu/ sets, even if the two extreme cases (/ɑdbɑ/, /ɑgbɑ/) are disregarded. This tendency (though not significant) is interesting since Repp (1979a) found shorter boundaries in stimuli of the type /V_1bgV_2/ when $V_1 = V_2$ than when $V_1 \neq V_2$. The $V_1 = V_2$ condition was met by the present /Cɑ/ set, since all VC stimuli began with /ɑ/. Thus, this difference might reflect a perceptual effect of contextual homogeneity, with a possible basis in articulation.

VC-CV combinations: C_1 responses and errors. To the extent that they do not derive from C_2 misidentifications, C_1 responses violate the principle that, at short silent intervals, C_2 is perceptually dominant over C_1. A high percentage of these responses occurred in /ɑdbu/ and /ɑgbu/; several subjects had difficulty perceiving the stop in /bu/ even at the longer silent intervals (cf. Repp, 1979a), most likely because this stimulus had only minimal formant transitions that were difficult to detect and therefore were overpowered by more pronounced cues in the preceding signal portion. C_1 responses were also frequent in /ɑbdi/, /ɑdbi/, and /ɑgbi/; they could only in part be accounted

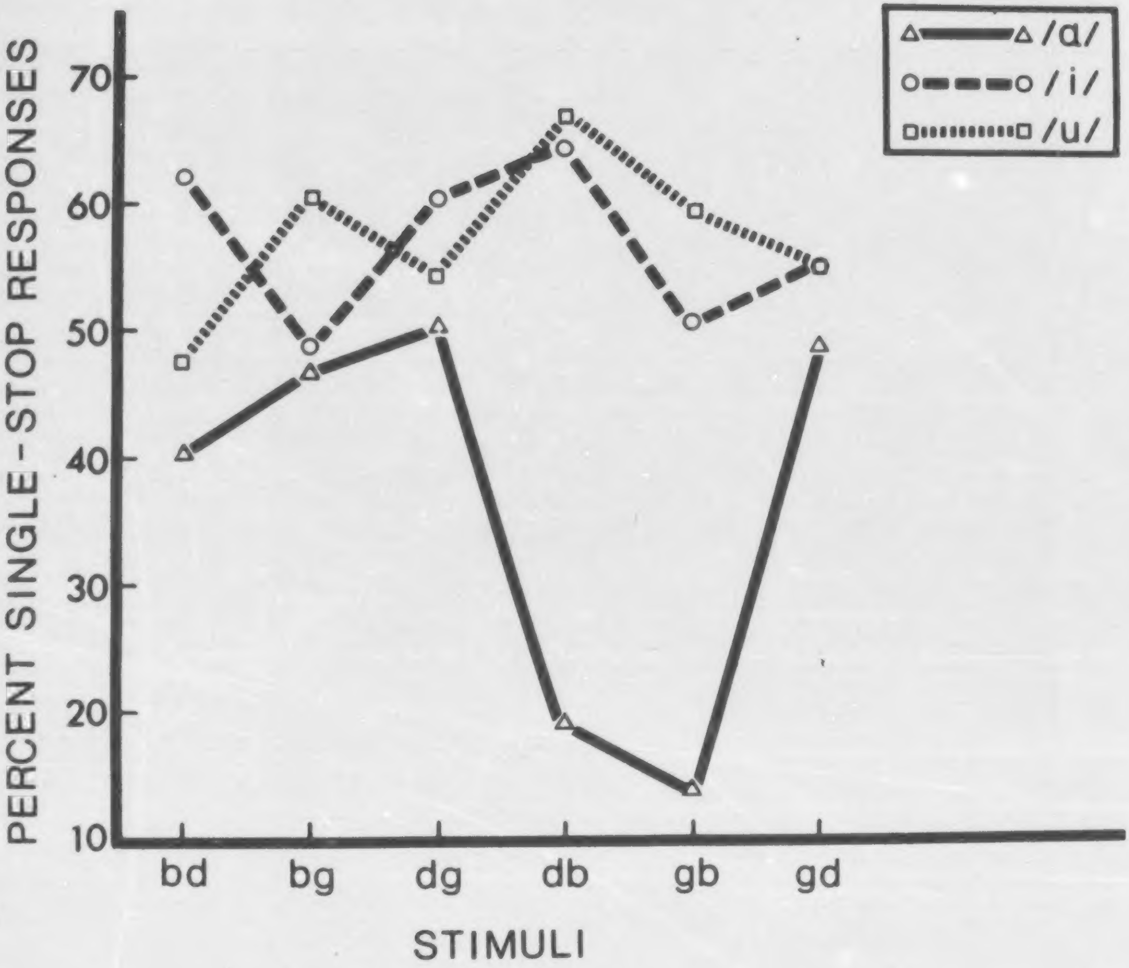


Figure 1. Percent single-stop responses (averaged over all silence durations) to the 18 VC_1-C_2V combinations (synthetic speech).

for by C_2 confusions between /bi/ and /di/. Many of the remaining C_1 responses could be predicted from the way the isolated stimulus components were perceived, except for a small percentage occurring in response to /ɑdbɑ/ and /ɑgbɑ/. Note that nearly all these cases involve labial stops in second position; thus, syllable-initial labial formant transitions seemed to be less effective in competition with conflicting syllable-final transitions than syllable-initial alveolar and velar transitions.

A large proportion of the error responses (responses other than the equivalents of C_1, C_2, and C_1C_2) could be predicted from the misidentifications of the monosyllabic components. There were certain unpredicted errors, however, that showed up with consistency. They included "bg" responses to /ɑdgɑ/ and especially to /ɑdgi/ (rarely to /ɑdgu/), which constituted the large majority of error responses to these stimuli (total: 9.2 percent); and "bd" responses to /ɑgdɑ/, /ɑgdi/, and /ɑgdu/, which made up about two thirds of the errors to these stimuli (total: 11.2 percent). These errors involve alveolar-velar combinations (in either order) in which the first stop was mislabeled as "b". (Neither /ɑd/ nor /ɑg/ was misidentified as "ab" in isolation.) We may be dealing here with a form of perceptual contrast (cf. Repp, 1978).

EXPERIMENT 2: PRODUCTION--ACOUSTIC MEASUREMENTS

Experiment 2 provided acoustic measurements of natural VC_1C_2V utterances, in order to see whether there is any relationship between the amount of silence required in perception and the average durations of closure periods in natural speech. While there have been several studies of closure durations associated with single intervocalic stops, the only study of two-stop sequences to date seems to be the unpublished work of Westbury (Note 1). However, he examined only clusters that were heterogeneous with respect to voicing (i.e., clusters of one voiced and one voiceless stop), whereas the present study was concerned with sequences of two voiced stops. Nevertheless, his results are highly relevant. He found that total closure durations were shorter when the first stop was alveolar than when it was labial or velar; they were also shorter when the second stop was velar than when it was alveolar or labial. In addition, he found an effect of vocalic environment, which he interpreted as a tendency towards temporal compensation for intrinsic variations in vowel duration: the longer the duration of the context (/bVC_1C_2Vt/), the shorter the closure duration. He did not report any changes in the effects of stop place of articulation across different vocalic environments.

The present study not only used somewhat different stimulus materials but also went beyond Westbury's by dividing closure periods into two portions. This was possible since most of the utterances measured contained release bursts of the syllable-final stop (C_1). (Westbury's utterances either did not contain such bursts, or he did not take them into account in his measurements.) In perceptual studies using natural speech, C_1 release bursts are deleted to produce the perceptual phenomenon of interest (Raphael & Dorman, in press; see Exp. 3 below). However, since the acoustic information for the syllable-final stop really includes the C_1 release and the preceding closure, this fact needs to be taken into account in any explanation of perceptual results: It may be that the amount of silence listeners need in perception is more directly related to the closure preceding the release ("C_1 closure") than to the total closure duration in production.

Method

Subjects. The subjects were two female research assistants, both native speakers of American English, and the author. The author, a native speaker of the Viennese variety of German, has lived in the United States for over 11 years but has retained a foreign accent. However, it was considered unlikely that the pronunciation of voiced stop consonant sequences in meaningless isolated disyllables would show any systematic influence of native language.

Utterances. The utterances were the same as in Experiment 1. The 18 disyllables were arranged into 10 different random lists that were typed onto a sheet of paper in simple spelling (e.g., abdi, adgu, etc.). After listening to sample pronunciations and practicing for a few minutes, the subjects read from the lists at an even pace, pronouncing each utterance at a fairly fast rate, with stress on the second syllable but without neutralizing the initial vowel. The recordings were made in a soundproof booth, using a Shure microphone and an Ampex AG-500 tape recorder.

Measurement procedure. All measurements were performed on a large-scale oscillographic display provided by a GT40 computer. After inputting an utterance from audio tape, critical points in its digitized waveform were located in the continuous, magnified display by means of a cursor, and the distance from one critical point to the next was measured to the nearest tenth of a millisecond using an automatic counter. Seven measurement points were defined:

A. Approximate onset of utterance.
B. Offset of VC portion. (Sometimes, voicing pulses persisted into the closure; in this case, the onset of significant damping--indicating closure of the vocal tract--was taken as the criterion.)
C. Onset of C_1 release burst.
D. Offset of C_1 release burst (approximate within a few msec).
E. Onset of CV portion.
F. Onset of periodicity in CV portion.
G. Approximate end of utterance.

From these measurement points, the following durations were derived:

F - A = Total utterance.
    B - A = VC portion.
    D - B = Total closure.
        C - B = "C_1 closure".
        D - C = C_1 release burst.
        E - D = "C_2 closure".
    G - E = CV portion.
        F - E = C_2 burst and aspiration.
        G - F = CV voiced portion.

All measurements were performed by a research assistant (a graduate student in phonetics) after thorough consultation with the author. Analyses of variance were performed on all measures of interest, with the factors Speakers, Vowels (three final vowels), and Consonants (six combinations). Since C_1 and C_2 were not orthogonal factors, their separate influences were examined in post-hoc (Newman-Keuls) tests comparing those six pairs of utterances that differed in one component only (C_1 effects: /bg/ vs. /dg/, /bd/ vs. /gd/, /db/ vs. /gb/; C_2 effects: /gb/ vs. /gd/, /db/ vs. /dg/, /bd/ vs. /bg/). The pooled within-cell variance (10 observations per cell, i.e., per utterance) was taken as the error term. Missing values, due to rare mispronunciations or acoustic anomalies, were replaced with the cell mean prior to analysis.

Results and Discussion

Total closure duration. The pattern of average closure durations as a function of consonant combinations and final vowels is shown in Figure 2, separately for each speaker. The grand average duration was 168 msec, with an average within-cell standard deviation of 15 msec. Statistical analysis revealed, first of all, a speaker effect, $F(2,486) = 262.3$, $p \ll .001$: BHR's closures were longer (188 msec, on the average) than DK's (162 msec) and SP's (154 msec). More interestingly, there was a highly significant vowel effect, $F(2,486) = 36.1$, $p \ll .001$: Closure durations were shorter for final /ɑ/ (160 msec) than for final /i/ (172 msec) and /u/ (172 msec). This effect was shown (on the average) by all three speakers and by each individual consonant combination; no statistical interaction involving the vowel effect approached significance. Finally, there was a significant consonant effect, $F(5,486) = 8.5$, $p < .001$, which did not interact with any other factor, despite (or because of) the considerable variability evident in Figure 2. The six consonant combinations were arranged as follows: /dg/ (161 msec), /bg/ (165 msec), /db/ (168 msec), /gd/ (170 msec), /gb/ (172 msec), /bd/ (173 msec). Newman-Keuls tests revealed one significant effect of the first stop (/d/ shorter than /b/, $p < .05$) and two significant effects of the second stop (/g/ shorter than /b/ and /d/, both $p < .01$), out of three comparisons in each case.

Certainly, these data provide no evidence for closures to be shorter in back-front sequences than in front-back sequences, or to be especially short in /ɑdbɑ/ and /ɑgbɑ/ (cf. Exp. 1). However, the results are in excellent agreement with Westbury's (Note 1) measurements, which showed closures to be shortest for alveolar stops in first position and for velar stops in second position. Westbury also found, in agreement with the present results, that closure durations were shortest in /ɑ-ɑ/ context, and he related this finding to the relatively long durations of these vocalic portions. We will return to this issue below.

C_1 closure. The C_1 closure measurements are shown in the left half of Figure 3. Since speaker SP did not consistently produce C_1 release bursts, her closure durations could not be broken down into components. Speakers BHR and DK, on the other hand, produced release bursts in all utterances. Their average C_1 closure lasted 74 msec, with an average within-cell standard deviation of 13 msec. C_1 closures were significantly longer in BHR's productions (80 msec) than in DK's (67 msec), $F(1,324) = 90.8$, $p \ll .001$, which parallels the difference in total closure durations reported above. Interestingly, there was no significant effect of the final vowel here,

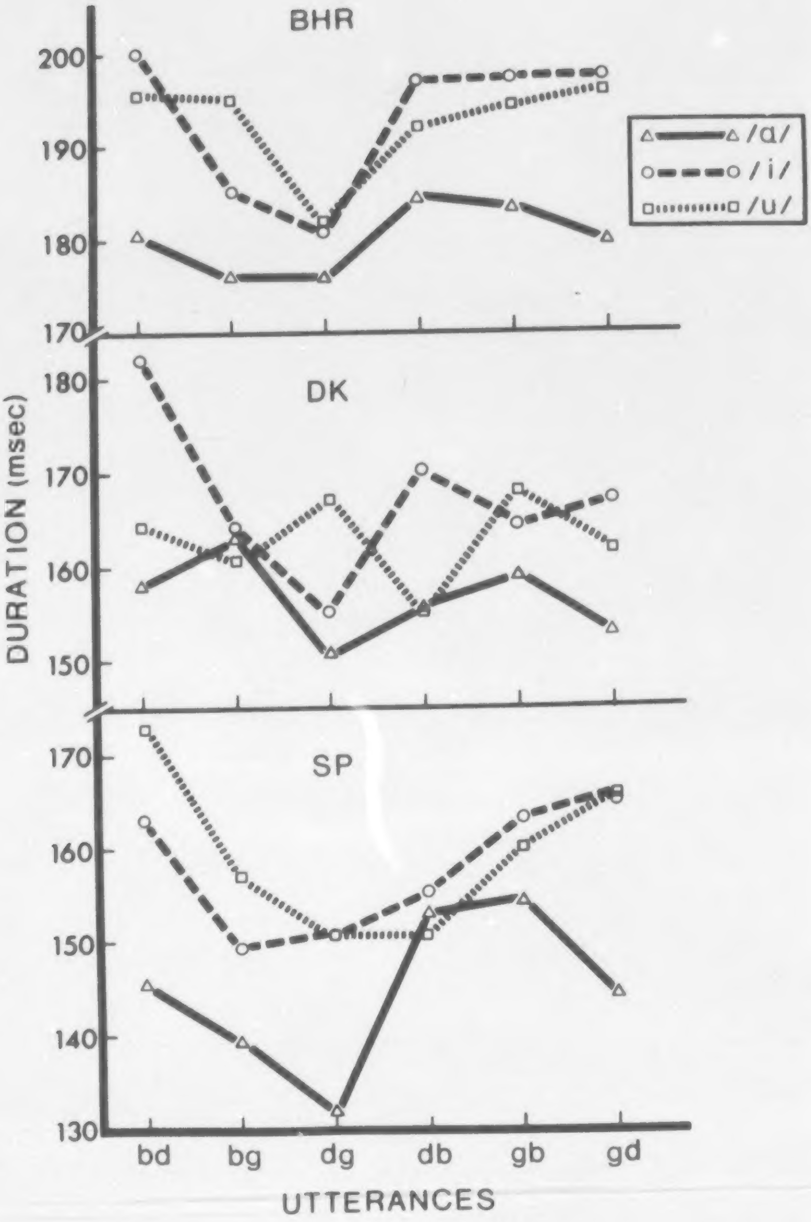


Figure 2. Average total closure durations in 18 VC_1-C_2V combinations produced by three speakers.

although there had been such an effect on total closure duration. However, there was a highly significant effect of consonants, $F(5,324) = 34.7$, $p < .001$, which also interacted with speakers, $F(1,324) = 3.7$, $p < .01$. Averaging over speakers (which seems permissible since the interaction was quite small), the rank order was: /gb/ (63 msec), /db/ (66 msec), /dg/ (70 msec), /gd/ (72 msec), /bg/ (80 msec), /bd/ (88 msec). Newman-Keuls tests showed C_1 closures to be clearly longer when C_1 was /b/ than when it was /d/ or /g/ ($p < .01$)--a result that is in striking agreement with measurements of closure durations in single intervocalic stops, which show longer durations for labials (e.g., Kohler, 1979; Umeda, 1977; Westbury, Note 1). However, C_2 also affected C_1 closure duration: C_1 closures were longer preceding /d/ than preceding /b/ ($p < .01$) or /g/ ($p < .01$, but only shown by speaker DK). Thus, while a syllable-final /b/ led to long C_1 closures, a following syllable-initial /b/ was associated with rather short C_1 closure durations.

C_2 closure. The C_2 closure measurements for speakers BHR and DK are shown in the right half of Figure 3. The average C_2 closure lasted 84 msec, with an average within-cell standard deviation of 16 msec. BHR's C_2 closures were significantly longer (90 msec) than DK's (78 msec), $F(1,324) = 46.1$, $p \ll .001$, as had been his C_1 closures. There was a significant vowel effect, $F(2,324) = 6.9$, $p < .001$, C_2 closures being shorter preceding /ɑ/ (80 msec) than preceding /i/ (85 msec) or /u/ (87 msec). Since C_1 closure had shown no vowel effect, it was C_2 closure that was responsible for the variations in total closure duration with final vowel. C_2 closure durations varied significantly across different consonant combinations, $F(5,324) = 13.2$, $p < .001$, and the pattern differed somewhat between the two speakers, $F(5,324) = 4.2$, $p < .001$. Overall, however, the rank order was nearly the inverse of that for C_1 closure duration: /bd/ (75 msec), /dg/ (78 msec), /bg/ (82 msec), /gd/ (84 msec), /db/ (90 msec), /gb/ (95 msec). Newman-Keuls tests showed that syllable-initial /b/ (C_2) was associated with longer C_2 closures than either /d/ or /g/ ($p < .01$), with somewhat longer closures for /g/ than for /d/ ($p < .05$), whereas C_2 closures were shorter when the preceding stop was /b/ than when it was /g/ ($p < .01$). Thus, C_2 closures, like C_1 closures, were longest when the associated stop was labial, but tended to be short when the *other* stop was labial.

Other signal portions. Since only closure duration measures are directly relevant to the topic of this paper, the other measurements will be summarized only very briefly. *C_1 release bursts* (average duration 17 msec) were markedly shorter for syllable-final /b/ in BHR's utterances, but not in DK's. *VC portions* (average duration 105 msec) showed no speaker difference (in contrast to the closure measures) but an effect of C_1: The vocalic portion was shorter for /b/ than for either /d/ or /g/ ($p < .01$). The *C_2 burst and aspiration portion*--the voice onset time (VOT) of C_2--showed the familiar effect of C_2 place of articulation, VOTs being shortest for /b/ (11 msec) and longest for /g/ (24 msec), with /d/ (19 msec) in between. Two speakers (BHR and DK) had shorter VOTs before /ɑ/; speaker SP, however, showed the opposite pattern. (SP also had much shorter VOTs than the other two speakers.) The *voiced CV portion* (average duration 221 msec) was longer for /ɑ/ for two speakers; again speaker SP differed by showing no vowel effect. There was no consonant effect here but a speaker difference, DK being slower than BHR (and both much slower than SP). Since DK had shorter closures than BHR, and since VC portions showed no speaker differences, independent temporal control of the different signal portions is suggested.

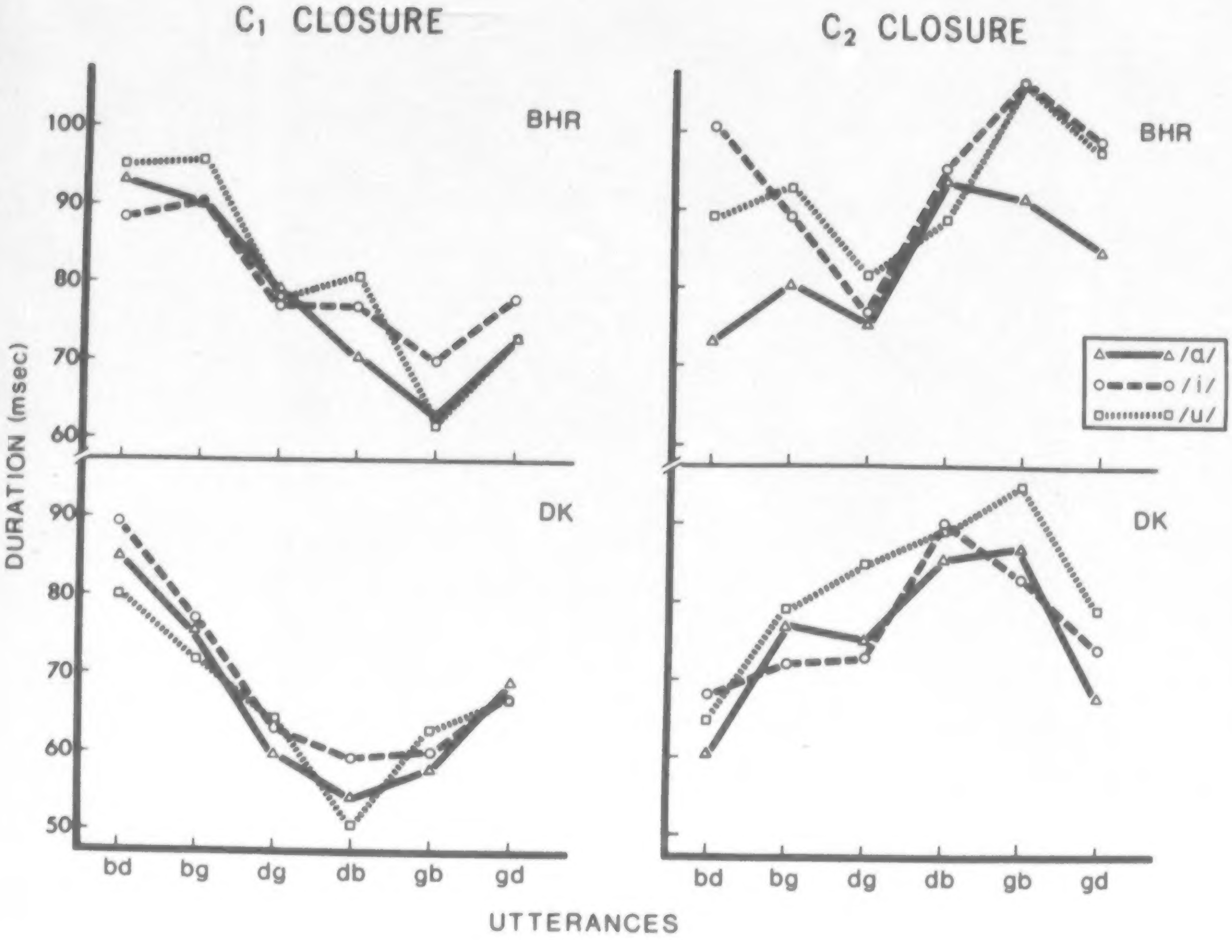


Figure 3. Average C_1 and C_2 closure durations in 18 VC_1-C_2V combinations produced by two speakers.

Summary. Closure durations were affected by the identity of both consonants as well as by the final vowel. C_1 and C_2 generally had opposite effects; thus, total closure durations ranked /d/ < /b/ < /g/ with respect to C_1 but /g/ < /b/ < /d/ with respect to C_2. When C_1 and C_2 closure segments were considered separately, however, the consonant effects were found to reflect primarily labial articulation: Both C_1 and C_2 closures were longest when the associated consonant was /b/ and tended to be shortened when the *other* stop was /b/. Total closure durations were shortest in /-ɑ/ context, and this effect was entirely due to variations in C_2 closure.

This pattern of results does not show a close resemblance to the perceptual results of Experiment 1. The abnormal perceptual boundaries for /ɑdbɑ/ and /ɑgbɑ/ have no parallel in production, and systematic effects of C_1 and C_2 across all three vocalic contexts are observed in production only, not in perception. Only the final-vowel effect (shorter closures in /-ɑ/ context) corresponds to a tendency towards shorter perceptual boundaries in that context. However, this effect could easily have an auditory basis: Several studies have shown that silent gaps are easier to detect in spectrally homogeneous than in heterogeneous environments (Collyer, 1974; Perrott & Williams, 1971; Williams & Perrott, 1972). Since the initial vowel in the present stimuli was always /ɑ/, stimuli ending in /-ɑ/ were spectrally more homogeneous than stimuli ending in /-i/ or /-u/, and perhaps this homogeneity facilitated the detection of the silent closure period.

EXPERIMENT 3: PERCEPTION--NATURAL SPEECH STIMULI

So far, our comparison of perception (Exp. 1) and production (Exp. 2) of two-stop sequences has been disappointing. However, the results of Experiment 1 may not have been representative, due to peculiarities of the synthetic stimuli. Although this possibility seems less likely in view of the good agreement between portions of the results of Experiment 1 and the earlier findings of Liberman (1975) and Raphael and Dorman (in press), it seemed desirable to replicate Experiment 1 using natural-speech stimuli. This was the purpose of Experiment 3.

Method

Subjects. Twelve subjects participated. They included ten paid volunteers with little experience in speech perception experiments and two subjects with considerable experience as listeners (a graduate research assistant and the author).

Stimuli. The stimuli were constructed from speaker BHR's utterances, which had been collected and measured in Experiment 2. To avoid token-specific irregularities and to permit an estimate of natural variability, four different tokens of each of the 18 utterances were selected from the 10 originally recorded. Thus, the initial stimulus pool consisted of 4 x 18 = 72 utterances. All utterances were digitized at 10 kHz and edited using the Haskins Laboratories Pulse Code Modulation system. The original closure periods (including the C_1 release bursts) were excised, and various amounts of silence (0-100 msec, in 10-msec steps) were inserted instead. The VC and CV portions were also stored in separate files.

The experimental tapes were analogous to those of Experiment 1. Three parallel sets were recorded, one for each final vowel. In each set, the first stimulus sequence consisted of the isolated VC and CV portions in random order, arranged in 5 blocks of 48, with ISIs of 2.5 sec and 10 sec between blocks. The 48 stimuli resulted from 4 tokens of each of 2 portions (VC and CV) of 6 utterances. The second stimulus sequence contained the VC-CV combinations in random order, arranged in 4 blocks of 66, with ISIs of 2.5 sec and 10 sec between blocks. The 66 stimuli resulted from 11 closure durations for one token of each of the six utterances. Different tokens were used in each of the four blocks; thus, there were in fact 4 x 66 = 264 physically different stimuli.

Procedure. Each subject participated in three sessions, one for each final-vowel condition. The order of final-vowel conditions was counterbalanced across subjects. In each session, the isolated VC and CV portions were presented first. A total of 5 responses for each token of each utterance was obtained, i.e., 20 responses for each utterance when token variation is ignored. Subsequently, the VC-CV combinations were presented three times, separated by appropriate rest periods. That is, each subject gave a total of three responses to each individual stimulus, or 12 responses when ignoring token variation.

Results and Discussion

Monosyllables. The natural-speech CV stimuli were quite intelligible, but the VC stimuli were less well identified than the synthetic stimuli in Experiment 1. The stop in /ɑg/, in particular, was frequently misidentified, with "b" confusions being about twice as frequent as "d" confusions. This poor identifiability was obviously a consequence of removing the C_1 release burst. The percentages of correct responses for the /Ci/, /Cɑ/, /Cu/, and /ɑC/ sets were 90.2, 96.3, 99.4, and 82.3, respectively (52.1 for /ɑg/). The confusion patterns did not seem to reflect in any way the context in which a given stimulus portion had been pronounced; thus, there seemed to be little coarticulation between VC and CV portions.

VC-CV combinations: Two-stop vs. one-stop responses. The results of the main part of the experiment were somewhat startling. Although the two experienced subjects produced what seemed to be typical and orderly results, a number of the naive subjects failed to show the VC-CV interference phenomenon, i.e., the predominance of single-stop percepts at short closure durations. *All* naive subjects reported two stops at short silence durations for at least some of the stimuli. Moreover, these responses were correct more often than not, and those misperceptions that occurred were typically consistent and stimulus-specific.

This outcome was quite unexpected, even though it will be recalled that two subjects in Experiment 1 had to be excluded for the same reason. To make sure that no problem of instructions was involved, two of the subjects were recalled and carefully instructed by the author. The same result was obtained: There were very few single-stop responses. Inspection of the stimuli did not reveal any reason for this "abnormal" behavior of the majority of listeners. Of course, researchers have known for a long time that speech cues--silence in particular--do not always have a perceptual effect: Their effect depends on the values of other relevant cues in the signal. In the present case, the formant transitions in and out of the closure and the C_2 release burst may have provided stop manner cues strong enough to override the perceptual effect of silence. What is surprising is that this occurred only for the naive listeners, as if they assigned less weight to the silence cue than the two experienced listeners. Interestingly, very similar observations have recently been reported by May, Porter, and Miller (1980).

Five subjects had to be excluded because they either gave no single-stop responses at all or just a few that were fairly randomly distributed. However, the responses of the remaining five naive listeners fell into a fairly orderly pattern that, moreover, resembled the results of the two experienced listeners, BHR and PP. Therefore, the data of all seven subjects were combined. They are plotted in Figure 4, which is analogous to Figure 1.

The figure first shows a pronounced vowel effect, $F(2,12) = 4.7$, $p < .05$: Considerably more silence was required to hear both stop consonants in /-i/ context than in /-ɑ/ and /-u/ contexts, and slightly less silence was required in /-ɑ/ context than in /-u/ context. While the latter tendency parallels the findings of Experiments 1 and 2, the first, larger difference has no correspondence in the earlier results. This difference was primarily due to the naive subjects since neither BHR nor PP showed any vowel effects that were consistent across all six consonant combinations. Inspection of the test schedule suggested that the effect was not an artifact of test order, which was still nearly balanced across the selected subjects.

The second effect seen in Figure 4 is a pattern of differences across the six VC-CV combinations, $F(5,30) = 8.1$, $p < .001$, that was quite consistent across the three final-vowel contexts. (The interaction was marginally significant.) In each case, the longest silences were required for /dg/; /bg/ ranked second in two contexts and third in the third. The shortest silence durations were required in /bd/ and /gd/, except in /-u/ context where /db/ had the shortest boundary. Once again, this pattern does not consistently follow the front-back vs. back-front distinction. Rather, it seems to reflect an effect of C_2: Longer silences were required when C_2 was /g/ than when it was either /b/ or /d/ ($p < .01$ in Newman-Keuls tests). Note that the boundary rank order /g/ > /b/ > /d/ with regard to C_2 is precisely the opposite of that obtained in production, indicating that VC-CV combinations with *longer* total closure durations in production required *less* silence in perception. This runs counter to the articulatory hypothesis, as conceived at the outset.

VC-CV combinations: C_1 responses and errors. Given the high frequency of /ɑg/ misidentifications, a large number of errors, as well as single-stop responses at long closure durations, might be expected in VC-CV stimuli containing that component. The errors did occur; however, single-stop responses were not as frequent as expected. To /gb/ combinations, subjects frequently responded "db"; and "bd" responses to /gd/ stimuli were extremely common. Thus, listeners tended to prefer that confusion of /ɑg/ that led to the perception of two stops over the one that led to single-stop responses, perhaps because of the acoustic inappropriateness of the /ɑg/ transitions for a single "b" or "d" percept. Other common confusions that could not be fully accounted for by misperception of the monosyllabic components were "gb" responses to /db/ and "bg" responses to /dg/. All these errors involved, of

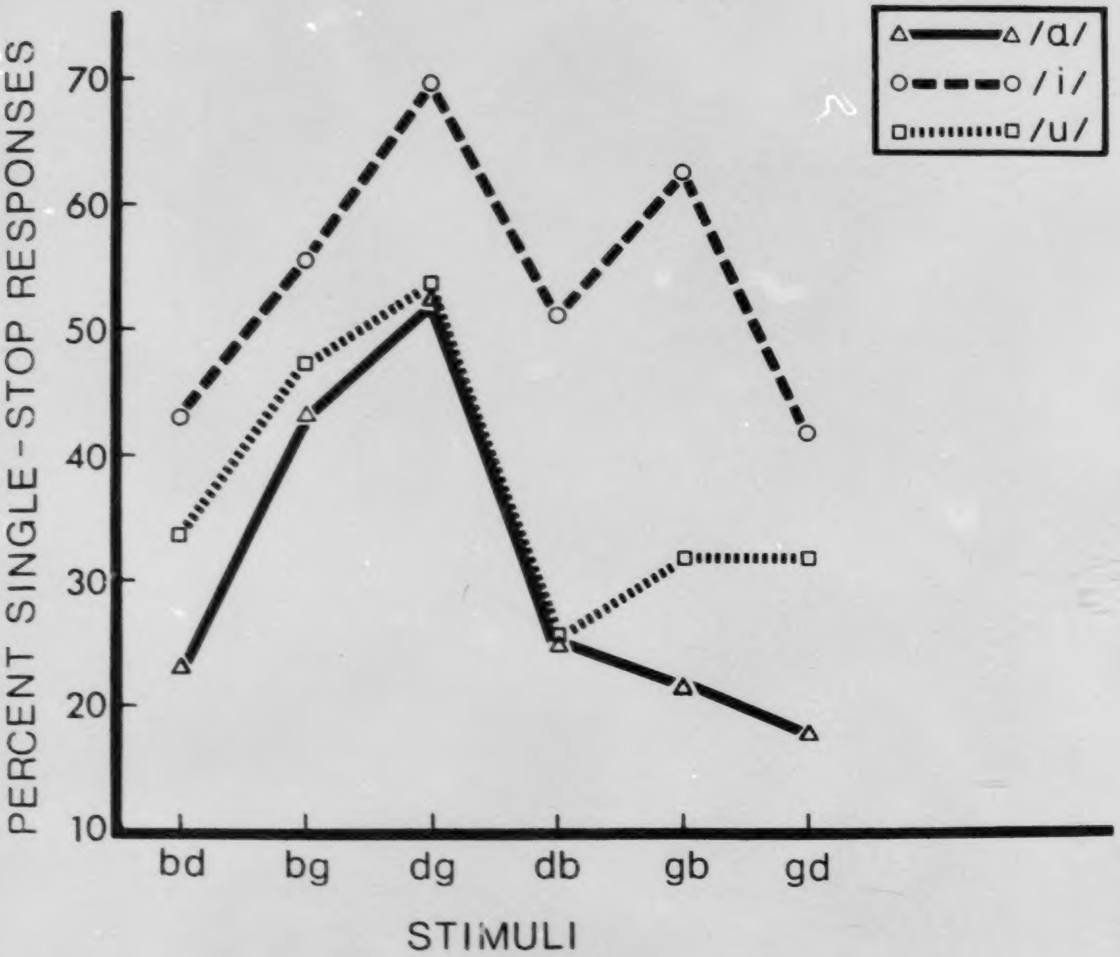


Figure 4. Percent single-stop responses (averaged over all silence durations) to the 18 VC_1-C_2V combinations (natural speech).

course, the perception of C_1; C_2 was very rarely misidentified. What is noteworthy is that a large number of errors occurred at all silence durations, including the longest, and that they were always in the direction of hearing two stops, rather than one. In other words, the listeners seemed to "know" that conflicting VC and CV cues could not be integrated into a single percept; it is not clear, however, what led them to misidentify C_1 so frequently in VCCV context. Note that the error pattern in the present experiment resembled that found in Experiment 1.

CONCLUSIONS

Systematic variations in the amount of silence required to hear two stops in utterances of the VC_1C_2V type do not appear to be correlated with variations in closure durations of corresponding natural utterances. They even differ a good deal between perceptual experiments employing synthetic and natural stimuli, respectively. Thus, the cause for the perceptual variability must be sought in auditory properties of the stimuli; it does not seem to be grounded in listeners' knowledge of articulatory dynamics. Presumably, the effective amount of silence perceived, or the effective value of some other relevant stimulus characteristic, is modified by the acoustic environment (in ways not yet understood) before it enters the phonetic decision process.

This conclusion underlines the importance of distinguishing between auditory and phonetic (or articulation-based) phenomena in speech perception. A number of perceptual effects have been reported that seem to require an explanation that makes reference to speech production (for recent examples, see Repp et al., 1978; Mann & Repp, 1980, in press). Indeed, the basic fact that silence plays a role at all in the perception of stop consonants may still belong in that category, although it also invites auditory hypotheses of various sorts. However, the present experiments, in conjunction with earlier data (Repp, 1979a, 1979b), suggest that variations in the amount of silence required for accurate perception arise at an auditory level. Since speech must pass through the auditory system on its way to higher centers of processing, we must expect that the perceptual phenomena we uncover in the laboratory will reflect both auditory and phonetic processes. To distinguish between these two sources of variation in each individual case is perhaps the most pervasive, and the most challenging, problem of speech perception research.

REFERENCE NOTE

1. Westbury, J. R. Temporal control of medial stop consonant clusters in English. Paper presented at the 93rd Meeting of the Acoustical Society of America in Austin, Texas, April 1977.

REFERENCES

Abbs, M. H. A study of cues for the identification of voiced stop consonants in intervocalic contexts. Unpublished doctoral dissertation, University of Wisconsin, 1971.

Collyer, C. E. The detection of a temporal gap between two disparate stimuli. Perception & Psychophysics, 1974, 16, 96-100.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 1979, 65, 1518-1532.

Kohler, K. J. Dimensions in the perception of fortis and lenis plosives. Phonetica, 1979, 36, 332-343.

Liberman, A. M. How abstract must a motor theory of speech perception be? Haskins Laboratories Status Report on Speech Research, 1975, SR-44, 1-15.

Lisker, L. Closure duration and the intervocalic voiced-voiceless distinction in English. Language, 1957, 33, 42-49.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. Perception & Psychophysics, 1980, 28, 213-228.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. Journal of the Acoustical Society of America, in press.

Massaro, D. W. Preperceptual images, processing time, and perceptual units in speech perception. In D. W. Massaro (Ed.), Understanding language. New York: Academic, 1975.

May, J. G., Porter, R. J., Jr., & Miller, C. J. Dichotic and monotic interactions between VC's and CV's: Masking or articulatory constraints? Journal of the Acoustical Society of America, 1980, 67 (Suppl. No. 1), S99-100.

Öhman, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, 1966, 39, 151-168.

Perrott, D. R., & Williams, K. N. Auditory temporal resolution: Gap detection as a function of interpulse frequency disparity. Psychonomic Science, 1971, 25, 73-74.

Raphael, L. J., & Dorman, M. F. Silence as a cue to the perception of syllable-initial and syllable-final stop consonants. Journal of Phonetics, in press.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. Perception & Psychophysics, 1978, 24, 471-485.

Repp, B. H. Influence of vocalic environment on perception of silence in speech. Haskins Laboratories Status Report on Speech Research, 1979, SR-57, 267-290. (a)

Repp, B. H. Influence of vocalic environment on perception of silence in speech: Amplitude effects. Haskins Laboratories Status Report on Speech Research, 1979, SR-59/60, 161-168. (b)

Repp, B. H. A range-frequency effect on perception of silence in speech. Haskins Laboratories Status Report on Speech Research, 1980, SR-61, 151-166.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 621-637.

Rudnicky, A. I., & Cole, R. A. Effect of subsequent context on syllable perception. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 638-647.

Umeda, N. Consonant duration in American English. Journal of the Acoustical Society of America, 1977, 61, 846-858.

Williams, K. N., & Perrott, D. R. Temporal resolution of tonal pulses. Journal of the Acoustical Society of America, 1972, 51, 644-647.

WORDS WRITTEN IN KANA ARE NAMED FASTER THAN THE SAME WORDS WRITTEN IN KANJI*

Laurie B. Feldman+ and M. T. Turvey+

Abstract. Two adult Japanese named colors written in Kanji, a logographic orthography, and in Kana, a syllabary. Although colors are more frequently written in the Kanji form and although Kanji are more compact graphic representations of words in general, latency to vocalization was consistently less for the Kana. This superiority is attributed to the closer relation of Kana to phonology and, therefore, to speech. The demonstrated greater facility for naming Kana accords with observations in the literature that very familiar visual configurations are consistently named faster when they conform to a phonographic principle than when they do not.

The evolution of writing systems is characterized by a trend away from representing many concrete morphological units towards representing a more restricted set of abstract phonological units. The characters of the oldest systems depicted objects and situations. These pictographs and semasiographs did not represent words. Their iconic quality made them visually distinctive, but they could refer to only a few concrete objects and common rituals. As these drawings became more conventionalized and their resemblance to specific objects diminished, the linguistic value of the character as the symbol for a spoken word was enhanced. Since a symbol could represent any word, logographs provided for expanded expression. For explicit written communication, however, a large number of characters had to be developed, usually according to a morphological principle. In Chinese, for example, semantically related words were often visually similar as they contained a common radical. Their particular pronunciation, however, was not specified in the written form. The subsequent introduction of phonology into orthography--phonetization (Gelb, 1952)--occurred at many levels. In rebus writing, words that sounded alike were represented by the same sign although their meanings were unrelated. These were substitutions for the whole word, but the same principle could be applied by syllable. The syllabary evolved from a logography and represented a deliberate and consistent use of a phonographic principle by which signs consistently represented the syllable. The Japanese syllable signs are derived from the Chinese logograms in this way. Later, in development of the alphabetic orthography, a further refinement of this principle occurred:

---

*Also in Language and Speech, 1980, 23, 141-147.

+Also University of Connecticut.

Acknowledgment: The research reported here was supported by NICHD Grant HD01994 awarded to the Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

Signs came to represent phonemes. By developing an orthography in which phonology is specified, more precise communication was possible with a reduced quantity of signs. It is apparent that the introduction of a phonologic principle renders an orthography more exact but its import to the reader is more equivocal.

The present study will investigate the role of orthographic structure in reading aloud. Baron (1977) delineates two plausible strategies by which naming or access to phonology can occur for an alphabetic script: an orthographic mechanism that uses letter-sound correspondences and focuses on component elements, and a word-specific mechanism that relies on larger visual patterns, either whole words, transgraphemic features or morphological units. The Japanese language is written in two scripts whose characteristics suggest this distinction by strategy. Of the two orthographies, only one is phonographic and would permit a (modified) orthographic mechanism. In Kana, a syllabary, the phonetic characterization of each syllable is represented by a character. By contrast in Kanji, a logography, each word is represented by one character such that no reliable description of pronunciation is available within the written form. With respect to Baron's (1977) distinction, naming in Kana, as in English, would seem to permit exploitation of either strategy, while naming in Kanji, because of its nonphonographic property, must entail a word-specific mechanism.

Baron's (1977) word specific mechanism can be interpreted as a lexical mediation of phonology. If naming a word occurs after lexical access, then naming latencies and lexical decision latencies should correlate since they both require lexical access. This hypothesis rests on the assumption either that a common lexicon supports naming and lexical decision or that there are two lexicons, one semantic and one phonologic, with an identical principle of organization. In fact, Forster and Chambers (1973) found that for English words naming and lexical decision times do correlate, especially for words of high frequency. Their conclusion was that lexical access mediates availability of a phonological code for naming. A general facilitation by frequency of occurrence has been demonstrated in many lexical tasks and is often incorporated into models of lexical organizations so that, for example, more frequent words should be named more quickly than less frequent words. If phonological structure is always derived by a lexical intermediary, then the value of a phonographic orthography is unclear and it is difficult to account for the results of Baron and Strawson (1976). These investigators showed that for skilled readers, latency to vocalization (naming) is faster for words that adhere to regular spelling-sound correspondences, for example, tone vs. gone or sweet vs. sword (Venezky, 1970), than for exception words that occur with greater frequency. This suggests the continued facilitation of a reliable sound-referencing or phonographic orthography for naming and implies that lexical access is not the only factor in latency to vocalization.

Brooks (1977) (and also Baron & Hodge, 1978) provides a similar demonstration of the effects of a phonology-referencing orthography. Using a small set of stimuli presented over several hundred trials, Brooks measured speed of naming. In the alphabetic condition, words were constructed from an artificial alphabet that adhered to a regular character-sound correspondence. They were compared with another condition in which the same responses were arbitrarily paired with the same visual configurations so that no functional alphabet obtained. While the arbitrary pairs were initially better, after practice the sound-correlated orthography proved superior in terms of shorter latencies to vocalization. When Brooks (1977) exaggerated the visual interaction within the forms by combining the component parts into a glyphic pattern, he found that this enhanced visual compactness also facilitated naming. In subsequent studies, he introduced controls both by expanding the stimulus vocabulary and by creating other artificial orthographies, but the reliance on contrived orthographies and extensive practice leaves lingering fears about the application of these results to skilled reading of natural orthographies.

The structure of the two writing systems in Japanese permits a natural language variation on the Brooks latency-to-vocalization procedure. Kana is a syllabary in which the phonetic specification of each syllable (more precisely mora) is depicted by a character. By virtue of this sound-referencing or phonographic orthography, similar sounding words look alike. In contrast, the Kanji script is logographic--there is no structure internal to the whole character that denotes pronunciation. Moreover, where Kana are generally used to designate tense, prepositions, new words and foreign terms, Kanji characters are used for nouns, verbs and adjectives. Finally, the Kanji tend to be compact and square, whereas the Kana tend to be a horizontal arrangement of discrete curved segments. By analogy with Brooks (1977), we compared latency to vocalization for Japanese color names written in Kana and in Kanji.

Phonographic writing systems specify the sounds of speech. Given the major outcome to Brooks' experiments, we should expect the latency of naming to be shorter for Kana than for Kanji. Against this expectation, however, are the following: First, Forster and Chambers (1973) demonstrated a strong positive correlation between the frequency of English words and naming time. Based on this evidence, we might suppose that because color names in Japanese literature appear more frequently in Kanji than in Kana, naming the colors written in Kanji should be faster than naming the colors written in Kana. Second, Brooks demonstrated, as noted above, that glyphic patterns were named more rapidly than their discrete counterparts. Therefore, we might expect shorter naming latencies for the somewhat glyphic Kanji forms than for the somewhat discrete Kana forms of the color names.

Procedure

Stimuli consisted of six Japanese color names whose English equivalents ranged in frequency from three to 203 occurrences based on the Kučera-Francis (1967) corpus of 50,000 word types. Each word had between two and four syllables when pronounced. Each color name occurred equally in its Kanji and its Kana form. Half of the Kanji were composed of two characters and half contained only one. See Table 1 for a summary of stimulus-item structure.

Table 1

Summary of stimulus-item structure

Japanese Words	English Equivalents	Number of Characters	Number of Syllables	Number of Strokes	Frequency of Occurrence
Kuro	black	1	2	11	203
Midori	green	1	3	14	116
Chairo	brown	2	3	9 + 6	176
Hairo	gray	2	4	6 + 6	80
Shuriro	vermilion	1	3	6	3
Kuriiro	chestnut	2	4	10 + 6	5

Two native Japanese served as subjects. They were instructed to read as rapidly as possible the stimulus words handwritten on slides displayed in two fields of a Scientific Prototype Model GB Tachistoscope. Each item was exposed for 500 msec and followed by a dark interval of about a second. The signal to light the display also triggered a timer that stopped at the onset of vocalization. In the course of three sessions, the two orthographic forms (Kanji/Kana) of the six color names were each presented 100 times in a randomized order.

In summary, the experimental design consisted of subjects' vocalizations of two orthographic forms (script) of each of six color names (stimulus items) presented in three sessions. Each session was composed of six trials per item where each trial was the average of approximately five observations, and data were then averaged over the six trials.

Results and Discussion

An analysis of variance pooled across all six stimulus items in each script condition for each subject revealed significant main effects for script, $F(1,10) = 66.88$, $p < .001$, session, $F(2,20) = 43.77$, $p < .001$, and subject, $F(1,10) = 25.02$, $p < .001$. The script x session interaction was significant, $F(2,20) = 8.48$, $p < .01$. As evident in Table 2, the facilitation

Table 2

Individual word latencies as a function of writing system and session

		Session I		Session II		Session III	
	Word	Kana	Kanji	Kana	Kanji	Kana	Kanji
1.	Kuro	458	470	423	440	409	424
2.	Midori	429	445	401	436	401	424
3.	Chairo	495	488	444	466	434	454
4.	Hairo	478	487	430	447	425	443
5.	Shuriro	488	507	460	480	443	468
6.	Kuriiro	532	539	456	486	468	501

of Kana relative to Kanji increases over sessions. The subject x session interaction was significant, $F(2,20) = 75.45$, $p < .001$.

When subjects' data were pooled, only script was significant, $F(1,1) = 192.15$, $p < .046$. Stimulus items approached significance, $F(5,5) = 4.48$, $p < .063$.

A significant facilitation of vocalization for the sound-referencing Kana orthography relative to the logographic Kanji orthography obtained for almost all stimulus words throughout all sessions. Naming latencies to the Kana averaged 18 msec faster than to the Kanji. (Any comparison of specific stimulus items must be made cautiously, as the acoustics of differing initial segments may have triggered the timer at different points in the utterance.) This result is impressive, as it violates documented effects of word structure related both to general usage, i.e., word frequency, and to visual scanning of discrete linear vs. compact glyphic patterns. By convention, Japanese color words are usually written in Kanji, but the familiarity of this form proved to be of no significant benefit. In addition, enhanced visual compactness, characterized by the square glyphic pattern and demonstrated by Brooks (1977) to be easier to scan than discrete linear forms (such as Kana), did not obscure the outcome. For latency to vocalization, Kana is faster than Kanji.

Japanese Kanji has been cited as an example of a script that does not contain information about phonology and recruited as evidence that readers must be able to access the lexicon visually in order to obtain a phonological specification. Another perspective on the same issue is the role of the lexicon in providing phonological codes for tasks such as naming. The structure of Kanji would seem to imply that such mediation is mandatory. In contrast, the lexical mediation of phonology may be optional in Kana, given its phonographic character.

At this point, it is perhaps useful to appreciate orthographic structure relevant to particular conditions in an attempt to account for the continued facilitation for reading aloud of Kana relative to Kanji. There is some developmental evidence that reflects this influence of orthographic structure on lexical performance. Steinberg and Yamada (1978) found that among three- and four-year-olds, the relative difficulty of learning Kana symbols far exceeded learning Kanji words. Sakamoto (in press) reports that while a small set of Kanji characters is systematically introduced by grade in the school curriculum, learning to read in Kana is completed in a relatively short period once the child begins to read.

Evidence of selective impairment and hemispheric superiority in word recognition also supports a distinction in processing the two Japanese orthographies. On both a visual recognition and a writing task (Sasanuma, 1974; Sasanuma & Fujimura, 1971), apraxic aphasics make more errors on the Kana than on the Kanji while simple aphasics perform comparably on Kanji, but make fewer errors on Kana. It seems that the Kana specification of phonology is not exploited by the apraxic. One interpretation (Sasanuma & Fujimura, 1971) is that the phonology-related pathology of the apraxic aphasic renders impossible the recognition of graphic forms as particular phonological patterns. Since Kana forms must be treated by the phonological processor in order to be identified, they are more vulnerable to left hemisphere damage than a Kanji transcription, which can be directly identified without any phonological interpretation. Tachistoscopic recognition by normals presents a different balance of hemispheric activity for Kana and for Kanji. Hatta (1977) reports a right hemisphere superiority for recognition of Kanji words that complements the Sasanuma, Itoh, Mori, and Kobayashi (1977) finding of left hemisphere superiority for Kana. A nonsignificant right hemisphere effect for Kanji (Sasanuma et al., 1977) may reflect differences in stimulus structure between these two experiments. Where Hatta used individual Kanji characters, Sasanuma et al. used random pairs of characters, but the combination of Kanji characters will often determine the semantic and phonological interpretation of each character (Martin, 1972).

Phonology is specified in the component elements of a Kana orthography such that the name of any previously unencountered words or nonwords may be generated; however, more specific experience with a particular character (or some combination of characters) is required to name Kanji. In some sense, there are more visual units to be considered by the orthographic mechanism for Kana than by the word-specific mechanisms for Kanji, but the redundancy of orthographic characters must get exploited in Kana. It is the sound-referencing or phonographic quality that permits the set of characters to be limited and generative.

These results represent an extension of the Brooks (1977) finding. The mora-sized graphemes of Kana are analogous to the phoneme-sized graphemes of an artificial alphabet. They both adhere to a phonographic principle. In a naming task, the advantage of a phonographic script relative to a logographic script is again manifest. To conclude, it seems that a delineation of strategies appropriate for a reading task such as naming must consider the particular properties of the writing system as well as the specific task, and that it is the specification of phonology intrinsic to its orthographic form that accounts for the facilitation of Kana relative to Kanji.

REFERENCES

Baron J. Mechanisms for pronouncing printed words: Use and acquisition. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, N.J.: L. Erlbaum, 1977.

Baron, J., & Hodge, J. Using spelling-sound correspondence without trying to learn them. *Visible Language*, 1978, *12*, 55-70.

Baron, J., & Strawson, C. Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 386-393.

Brooks, L. Visual pattern in fluent word identification. In A. Reber & D. Scarborough (Eds.), *Toward a psychology of reading: The Proceedings of the CUNY conferences*. Hillsdale, N.J.: Erlbaum, 1977.

Forster, K., & Chambers, S. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 627-635.

Gelb, I. *A study of writing*. Chicago: University of Chicago, 1952.

Hatta, T. Note. Recognition of Japanese Kanji in the left and right visual fields. *Neuropsychologia*, 1977, *15*, 685-688.

Kučera, H., & Francis, W. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press, 1967.

Marcel, A., & Patterson, K. Word recognition and production: Reciprocity in clinical and normal studies. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: L. Erlbaum, 1978.

Martin, S. Nonalphabetic writing systems: Some observations. In J. F. Kavanagh, & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, Mass.: MIT Press, 1972.

Patterson, K. E., & Marcel, A. J. Aphasia, dyslexia and the phonological coding of written words. *Quarterly Journal of Experimental Psychology*, 1977, *29*, 307-318.

Sakamoto, T. Reading of Hirangana. In R. Venezky (Ed.), *Orthography, reading and dyslexia*. Baltimore, Md.: University Park Press, in press.

Sasanuma, S. Kana and Kanji processing in Japanese aphasic. *Brain and Language*, 1975, *2*, 369-383.

Sasanuma, S. Kanji versus Kana processing in alexia with transient agraphia: A case report. *Cortex*, 1974, *10*, 89-97.

Sasanuma, S., & Fujimura, O. Selective impairment of phonetic and non-phonetic transcription of words in Japanese aphasic patients: Kana vs. Kanji in visual recogition and writing. *Cortex*, 1971, *7*, 1-18.

Sasanuma, S., Itoh, M., Mori, K., & Kobayashi, Y. Tachistoscopic recognition of Kana and Kanji words. *Neuropsychologia*, 1977, *15*, 547-553.

Steinberg, D., & Yamada, J. Are whole word Kanji easier to learn than syllable Kana? *Reading Research Quarterly*, 1978-79, *1*, 88-99.

Venezky, R. *The structure of English orthography*. The Hague: Mouton, 1970.

SOME EXPERIMENTS ON THE ROMAN AND CYRILLIC ALPHABETS OF SERBO-CROATIAN*

G. Lukatela+ and M. T. Turvey++

CROSS-LANGUAGE COMPARISONS: SERBO-CROATIAN ORTHOGRAPHIES AND THEIR SPECIAL PROPERTIES

Much if not most of current theorizing on the reading process and visual information processing is based on investigations with English language materials. Perhaps such processes vary but little across languages and orthographies and therefore a theory based on one language will suffice for all. However, what variations there are may prove to be revealing. We have been asking whether or not the reading of Serbo-Croatian may make use of different characteristics of the written word or different encoding routines than are used in the reading of English.

A distinction that is often made between logographic writing systems, such as Korean, Chinese, and Japanese kanji, and alphabetic systems, such as English and Serbo-Croatian, is that the former refer to the morphology, while the latter refer to the phonology. The logographic system is said to specify units of meaning, whereas the alphabetic system is said to specify the sounds of the spoken language, although the distinction is not as sharp. Indeed, this interpretation of the alphabet is less than ideal as far as English is concerned, for the correspondence between written and spoken English is opaque: graphemes can be made silent by context and, in general, graphemes take on different phonetic trappings in different graphemic contexts. Looking for regularity in the English orthography, Gibson, Pick, Osser, and Hammond (1962) advanced the idea of a spelling pattern, a cluster of letters that corresponds to a sound. While individual letters in English do not have invariant phonemic interpretations, certain arrangements of letters do, particularly when their locations within words are taken into consideration. Whether or not the notion of spelling pattern is valid, the point is obvious: the cipher relating script to utterance in English is complex. We argue that the cipher in Serbo-Croatian is considerably more transparent; and that for the Serbo-Croatian orthography the claim that it specifies the sounds of speech is potentially closer to the mark. But let us pursue the English orthography a little further.

---

*Also in Orthography, Reading, and Dyslexia, edited by James F. Kavanagh and Richard L. Venezky. Baltimore: University Park Press, 1980, pp. 227-247.

+University of Belgrade.

++Also University of Connecticut.

Acknowledgment. This research was supported in part by NICHD Grant HD-08495 to the University of Belgrade, and in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

The opaqueness of the script to utterance relation in English is owing, by and large, to two reasons. First, the pronunciation of the language evolved along different lines from the spelling of the language. Consider the following example cited by Henderson (1977). The English digraph *gh* as in *bough* and *rough* specified a unique guttural utterance until the seventeenth century. After the seventeenth century the pronunciation of *gh* took two directions: it either became silent as in *night* or took the phonemic interpretation /f/ as in *enough*. But the spelling had already become standardized largely owing to the efforts of the fifteenth century English printers, such as Caxton; and, in consequence, *gh* is handed down to the contemporary reader of English as an orthographic anomaly.

The second reason for the spelling-sound opaqueness is that the English orthography may be as close to the morphology as it is to the phonology. Indeed, in the evolution of the English language, Henderson (1977) has stated that the tendency has been for the orthography to reflect etymology, which is tantamount to saying that it reflects the basic units of meaning. In this vein Chomsky (1970) has argued that the English orthography is near optimal for writing the English language. The orthography preserves the morphology, which would not be the case if the optimality principles were phonemic correspondences. Thus, the spelling preserves the following morphological similarities--tele-graphy, tele-graph-ic, tele-graphy-y--in the face of the obvious phonetic variability. Similarly *anxious* and *anxiety* by virtue of their visual likeness permit the reader, in principle, to go directly from the appearance of the letter sequence to its meaning. Therefore, the fundamental point made by Chomsky (1970) and also by Venezky (1970) (but for somewhat different reasons) should be noted, namely, that the English orthography is systematic in its own right. It is specific to linguistic structure at a deep level and is not to be understood just as a phonemic transcription. Indeed, on the Chomsky-Venezky view, the script-utterance relation is opaque precisely because the script and utterance are *alternative* specifications of the same underlying structure (cf. Francis, 1970). However, the tempering conclusion of Gleitman and Rozin's (1977) thorough analysis is that it is not so much that English orthography is optimal for this or that grain-size of linguistic analysis, but rather that English writing is a rich mixture of a number of grains of linguistic representation, together with more than a sprinkling of arbitrary features.

Let us now turn to Serbo-Croatian, Yugoslavia's major language. Serbo-Croatian, unlike English, is pronounced as it is written; that is, individual letters have phonemic interpretations that remain consistent throughout changes in the context in which they are imbedded. All written letters are pronounced; hence, in Serbo-Croatian there are no silent letters and no double letters.

This state of affairs--a straightforward regularity between script and utterance--is by virtue of a historical development that sharply contrasts the evolution of the Serbo-Croatian orthography with that of the English orthography. The modern Serbo-Croatian orthography was constructed at the beginning of the nineteenth century by Karadžić on the basis of a simple rule: "Write as you speak and read as it is written!" In Serbo-Croatian, therefore, constraints on sound sequences are the sole sources of constraints on letter sequences. This contrasts with English in which restrictions on letter sequences derive not only from phonological constraints but also from a desire to preserve the etymology and graphemic conventions. That is, from a "...1400-year accumulation of scribal practices, printing conventions, lexicographers' selections, and occasional accident which somehow became codified as part of the present orthographic system" (Venezky & Massaro, 1979, p. 25). In English, illegal phonological sequences (such as /wh/) can be orthographically regular spellings (such as wh) but no such peculiarity is permitted in Serbo-Croatian.

Karadžić (1814) selected the speech spoken in mid-Yugoslavia as the ideal and to each phonemic segment of the speech he assigned a letter character or, in a few cases, a combination of letters. Karadžić took the majority of letters from the alphabet existing at the time but since the number of letters available was less than the number of phonemes needed, he borrowed or modified several letters from other alphabets. In fact, two alphabets were constructed: a Roman alphabet and a Cyrillic alphabet. In modern Yugoslavia, Eastern Serbo-Croatian uses primarily the Cyrillic script whereas Western Serbo-Croatian uses primarily the Roman. In some regions (e.g., Bosnia, Herzegovinia), however, both scripts are used about equally.

The Serbo-Croatian language has 30 phonemes. In the Cyrillic alphabet there is one letter for each phoneme; in the Roman, 27 phonemes are represented by single letters and three phonemes by pairs of letters: LJ, NJ, DŽ. Figure 1 compares the Roman and Cyrillic alphabets in uppercase and in Table 1 the letters (both uppercase and cursive) of the alphabets are given their corresponding letter-names in the International Phonetic Alphabet (IPA) transcription.

An important fact about the Roman and Cyrillic alphabets is that they map onto the same set of phones but still comprise two sets of letters that are, with certain exceptions, mutually exclusive. Of the total set of letters comprising the two alphabets the majority are unique to one or the other alphabet (see Figure 1). A number of letters, however, are shared by the two alphabets. Of these shared letters, some receive the same phonemic interpretation whether read as Roman or Cyrillic (referred to as common letters) and some receive two phonemic interpretations, one in the Roman reading and one in the Cyrillic reading (referred to as ambiguous letters). Therefore, one may recognize instances in which letters are different in shape but pronounced the same way, e.g., the Cyrillic И and the Roman I are both pronounced like the ea in seat; instances in which letters are the same in shape and pronunciation; and instances in which the letters are of the same shape but pronounced differently, e.g., the Cyrillic H is pronounced like the n in wine, the Roman H like the ch in the Scottish rendering of loch.

Three examples underscore the unusualness of Serbo-Croatian bi-alphabetism. The sentence, This is my mother, translated into Serbo-Croatian is spelled: TO JE MOJA MAJKA. In IPA it is rendered as: [to je moja majka]. There is no way to tell whether this particular sentence is written in Roman or Cyrillic, since only the common letters have been used. The sentence, The deer climbs, translated into Serbo-Croatian is spelled in Cyrillic as: CPHA CE BEPE. In IPA it is rendered as: [srna se vere]. However, if CPHA CE BEPE were read as Roman, it would be uttered as: [tspxa tse bepe], which is a meaningless utterance. Finally, one may note the sentence, The pupil studies

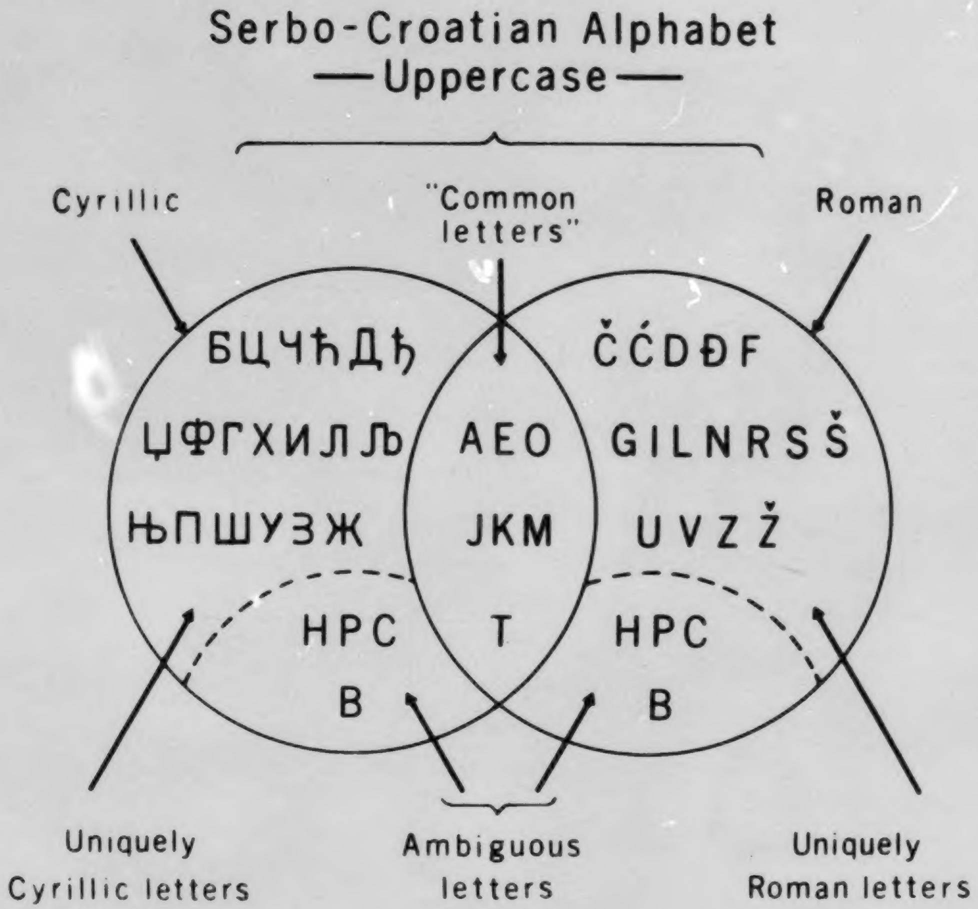

Serbo-Croatian Alphabet
—Uppercase—
Cyrillic
"Common letters"
Roman
БЦЧЋДЂ
ЏФГХИЛЉ
ЊПШУЗЖ
AEO
JKM
T
ČĆDĐF
GILNRSŠ
UVZŽ
HPC
B
HPC
B
Uniquely Cyrillic letters
Ambiguous letters
Uniquely Roman letters

Figure 1. The two alphabets of the Serbo-Croatian language.

Table 1. Letters of the Serbo-Croatian alphabet.

Serbo-Croatian				
Roman		Cyrillic		
Printed Upper Case	Cursive Lower Case	Printed Upper Case	Cursive Lower Case	Letter Name in IPA
A	*a*	А	*а*	a
B	*b*	Б	*б*	bə
C	*c*	Ц	*ц*	tsə
Č	*č*	Ч	*ч*	tʃə
Ć	*ć*	Ћ	*ћ*	tʃjə
D	*d*	Д	*д*	də
Đ	*đ*	Ђ	*ђ*	dʒjə
DŽ	*dž*	Џ	*џ*	dʒə
E	*e*	Е	*е*	ɛ
F	*f*	Ф	*ф*	fə
G	*g*	Г	*г*	gə
H	*h*	Х	*х*	xə
I	*i*	И	*и*	i
J	*j*	Ј	*ј*	jə
K	*k*	К	*к*	kə
L	*l*	Л	*л*	lə
LJ	*lj*	Љ	*љ*	ljə
M	*m*	М	*м*	mə
N	*n*	Н	*н*	ne
NJ	*nj*	Њ	*њ*	njə
O	*o*	О	*о*	ɔ
P	*p*	П	*п*	pə
R	*r*	Р	*р*	rə
S	*s*	С	*с*	sə
Š	*š*	Ш	*ш*	ʃə
T	*t*	Т	*т*	tə
U	*u*	У	*у*	u
V	*v*	В	*в*	və
Z	*z*	З	*з*	zə
Ž	*ž*	Ж	*ж*	ʒə

reading, which is written in Cyrillic ЂАК УЧИ ДА ЧИТА but in Roman as, ĐAK UČI DA ČITA. Regardless of which alphabet has been used, the phonetic transcription is the same in both cases: [dzjak uči da čita], as is the meaning.

A most central feature is that both alphabets are taught in the schools and by most accounts the letter forms and the letter-to-sound correspondences of both alphabets are learned by the end of the second grade. The children are taught one alphabet in the first year and a half and then master the other by the end of the second year. In the western part of the nation the Roman alphabet is learned first and in the eastern part of the nation it is the Cyrillic alphabet that the children master initially. This geographically based ordering of acquisition of the two alphabets provides a model for examining the relation of two separate symbol systems, learned at different times--a bi-alphabetism if you wish--of which bilingualism is the fashionable example. It deserves reemphasizing that the two alphabets map onto the same phonemic and semantic structure.

At this juncture let us collect the preceding discussions of the phonemic regularity and the bi-alphabetism of Serbo-Croatian in order to highlight several important contrasts with English orthography. First, where it can be claimed that the English orthography more directly represents the morphology, it can be claimed that the Serbo-Croatian orthographies more directly represent the phonology. Common to the views of Chomsky and Venezky, a reader of English often needs to know more about a word than its surface orthographic structure in order to pronounce it. One would say of Serbo-Croatian that knowledge about any word's surface orthographic structure is generally all that is needed in order to pronounce it. Second, English spelling more than occasionally reveals the etymology of words but the radical reworking of the Serbo-Croatian writing system according to Karadžić's injunction ensured that the contemporary orthography would be essentially ahistorical. Third, because of the virtually invariant relation between letter and sound there are no true homophones in Serbo-Croatian. (Situations such as *tale/tail*, *crews/cruise*, *wait/weight* could never arise.) We emphasize *true* because the bi-alphabetic nature of Serbo-Croatian permits homophones of a very special kind, precisely, letter sequences that are visually quite distinct--for one is composed mainly of uniquely Cyrillic and the other of uniquely Roman letters--but which are identical in pronunciation and meaning.

It is the case, however, that Serbo-Croatian, like English, allows true homographs. It is for this reason that a reader can *generally*, rather than *always*, pronounce a word correctly on the basis of knowing only its surface orthography. Two words may be written the same way, but, owing to different assignments of vowel length and accent type, can be pronounced differently and mean different things. In Serbo-Croatian a vowel can be short or long and its accent can or can not extend into the following syllable. Sometimes these contrasts are noted by diacritical marks. More commonly, however, the ambiguity must be resolved, as in English, by sentential context. The language gives rise additionally to a special kind of homography, again made manifest over the two alphabets. Thus a given letter sequence such as POTOP can be read one way in Roman and another way in Cyrillic (see Table 1), and mean two entirely different things (respectively, *inundation* and *rotor*).

There is a further feature of the Serbo-Croatian language on which we now pass remark by way of concluding our delineation of the language's special properties. It is that inflection is the principal grammatical device in the language in contrast with English, which uses inflection for grammatical purposes only sparingly. Thus for nouns, all grammatical cases in Serbo-Croatian are formed by adding to the root form an inflectional element, namely, a suffix consisting of one syllable of the vowel or vowel-consonant type. The Serbo-Croatian nouns, pronouns, and adjectives are declined in seven cases of singular and seven cases of plural whereas verbs are conjugated by person and number in six forms.

ERROR PATTERN IN BEGINNING READING

Where other languages with a close match between sound and writing have been examined, the evidence is that children learned very rapidly to read aloud letter sequences congruent with the orthographic rules of the language (Elkonin, 1963; Venezky, 1973). Nevertheless, it can be noted that indifferent to the script-to-utterance correspondence reading differences emerge early (Gibson & Levin, 1975) and that some children will continue to have problems even where the spelling of the words on which they are instructed is phonetically regular and maps to sound directly (Savin, 1972). Reading skill, in the long run, appears to be largely indifferent to the language being read (Gray, 1956). A not overly venturesome claim is that different writing systems induce differences in acquisition of reading and differences in the reading process without necessarily affecting the ultimate proficiency of reading. The point to be emphasized, perhaps, is that of Carroll (1972): "A perfectly regular alphabetic system may facilitate word-recognition processes but its use does not alter the fact that the learning of reading entails the acquisition of skills in composing word units from their separate graphic components and practice, large amounts of it, in recognizing particular word units."

Given the orthographic distinction between English and Serbo-Croatian one can ask: In what ways does the beginning reader in Serbo-Croatian differ from his counterpart in English and in what ways are they the same? One can ask, in short, with respect to the acquisition of reading, what changes across orthographies and what remains invariant? We are examining this question in relation to research already conducted and currently underway at the Haskins Laboratories.

A point of departure for the reading research of the Haskins Laboratories' group is that reading is somehow parasitic on speech. One recent focus has been the notion of "linguistic awareness" (Mattingly, 1972). A child might try to read words by the mediary of shape. But this nonanalytic strategy, while useful to a point, is far from optimal; the child cannot benefit from the fact that the alphabet permits its users to generate a letter string's pronunciation from the spelling. But what is required of the child to know how the alphabet works? I. Y. Liberman and Shankweiler (1979) argue that the child must realize that speech can be segmented into phonemes and he must know how many phonemes any given word in his vocabulary contains and their order. He must know that the letters of the alphabet represent phonemes, not syllables or some other unit of speech (see also Gleitman & Rozin, 1977; Rozin & Gleitman, 1977).

The difficulty and significance of phonemic segmentation has been frequently noted (e.g., Elkonin, 1973; Gibson & Levin, 1975; Rosner & Simon, 1971); the inability to analyze syllables into phonemes marks the child who has failed to learn how to read or, at least, who reads poorly (I. Y. Liberman, Shankweiler, A. M. Liberman, Fowler, & Fischer, 1977; Savin, 1972).

Exemplary of the difficulty with phonemic segmentation is the pattern of errors a child makes in reading syllables. For simple English consonant-vowel-consonant structures the error rate on the final consonant is larger than that on the initial consonant while the error rate on the vowel is largest of all (Shankweiler & I. Y. Liberman, 1972). Moreover, the form of the vowel and consonant errors differ in nontrivial ways (I. Y. Liberman & Shankweiler, 1979). To what extent, one might ask, are these patternings of errors orthographically based? Are they indigenous to the writing system of English or would they be as likely in the orthographies of Serbo-Croatian? For example, the greater error rate on vowels might be owing to the fact that in English vowel pronunciation is extremely context conditioned. On the other hand, it might be owing to the differential status of vowels and consonants in the perception and production of speech; in which case one might treat the different error rates of vowels and consonants and the direction of the difference as indexing a universal property of phonographic writing systems.

We have begun an examination of these questions through an experiment that is closely comparable to one previously conceived and conducted by the Haskins Laboratories group.

The 65 subjects in the experiment all tested within the normal range of intelligence. They were selected from the first grade population of an elementary school system located in Belgrade. Their ages ranged from 6.5 to 7.5 years. They had completed their first semester and had an active knowledge of the Cyrillic alphabet.

We devised two lists of the CVC-type monosyllables written in Cyrillic. One hundred CVCs were words and 100 CVCs were pseudowords. The words were familiar to first graders. In the word and pseudoword lists the 25 Serbo-Croatian consonant phonemes that can occur in both the initial and in the final positions of a word appeared twice in each position. In the majority of the trigrams the medial letter was one of the five Serbo-Croatian vowels (/i/, /e/, /a/, /o/, /u/) as in ДИВ 'giant,' ЦЕВ 'pipe,' ДАР 'gift,' СОК 'juice,' and ВУК 'wolf.' In some trigrams, however, the medial letter was the semi-vowel /r/. In Serbo-Croatian monosyllabic words of the type consonant-semivowel /r/-consonant, as in ВРХ 'top,' ТРН 'thorn,' ГРБ 'emblem,' are not infrequent. And finally, it should be noted that of the 100 words, 25 could be reversed to produce other words: For example the word БОР 'pine' if read from right to left reads РОБ 'slave.'

A string of three uppercase Cyrillic letters arranged horizontally at the center of a separate 3" x 5" white card defined a stimulus. The cards were placed face down in front of the subject and were turned over one by one by the examiner. The subject was asked to read each letter string aloud as it was presented. Responses were written down by the examiner and were recorded simultaneously on magnetic tape. A complete list was presented in a single session with each child participating in two separate sessions. If in the first session the child read the word list, then in the second session he read the pseudoword list and vice versa. The order of presentation was balanced across children.

The responses to the stimuli revealed several types of errors: 1) substitution, 2) addition, 3) omission, and 4) reversal of sequence when a letter string or a part of it was read from left to right. Single letter orientation errors did not occur because the Cyrillic uppercase letters did not provide opportunity for reversing letter orientation.

The analysis of errors showed that sequence reversals accounted for only a small proportion of the total of misread letters, although the lists were constructed to provide ample opportunity for the complete reversal of sequences. (As noted, 25% of the words were "reversible"; and 13% of the pseudowords were words if read from right to left, for example, the pseudoword НИС would become СИН 'son').

The complete sequence reversals are distinguished from the partial and the total reversal scores for words and pseudowords are given in Table 2. Proportions of opportunity for error (in percentages) are presented within parentheses. We note that sequence reversals were rare.

Single letter omission errors were also quite rare. Their distribution on initial and final consonants and on the medial vowel/semivowel is presented in Table 3. Omissions of the final consonant in words seem to be more frequent than in pseudowords, but the respective proportions of opportunity are too small to allow any reliable conclusion on their distribution.

Additional errors were distributed in a nonrandom manner (see Table 4). Additions of a single phoneme in front of the final consonant (FC_1) were more frequent than after the final consonant (FC_2), other types of additions being relatively infrequent.

In words and pseudowords of the consonant-semivowel /r/-consonant type, additions of a single phoneme in front of the final consonant were relatively the most frequent. For example, the word ГРБ was often misread as /grab/, /grub/, /greb/, or /grob/. In four words (ГРБ, ВРХ, ТРГ, ТРН) there were 45 single vowel additions, and in four pseudowords (БРС, ДРН, КРП, ПРК) there were 47 single vowel additions of the FC_1 type. Viewed in terms of opportunities for this particular error in the four words, the percentage amounts to 17% and in the four pseudowords up to 18%. This is a notable result. Apparently, to facilitate the phonetic rendition of the letter string, the child inserted a vowel between the medial semivowel and the final consonant.

Substitutions of single phonemes were the major source of errors in the experiment. Distribution of substitution errors on initial and final consonant and on the medial vowel/semivowel is presented in Table 5. Raw error

Table 2

Sequence reversals

	Complete sequence reversal	Partial sequence reversal	Total
Words	17% (1.1)	6 (0.0%)	23
Pseudowords	21 (2.5%)	13 (0.0%)	34

Table 3

Omission errors

	Initial consonant	Medial vowel	Final consonant	Total
Words	1	4	11 (0.2%)	16
Pseudowords	4	3	3	10

Table 4

Additions of a single phoneme

	Initial consonant	Medial vowel	Before final consonant FC1	After final consonant FC2	Total
Words	6	10	52	12	80
Pseudowords	1	9	52	25	87

Table 5

Single phoneme substitution errors

	Initial consonant	Medial vowel	Final consonant	Total
Words	172 (2.6%)	93 (1.4%)	264 (4.1%)	529
Pseudowords	213 (3.3%)	113 (1.7%)	368 (5.7%)	693

scores and the respective percentages (within parentheses) indicate that final consonant (FC) errors exceed initial consonant (IC) errors. A Wilcoxon signed-rank test on proportions of correct responses revealed that this difference was significant ($T_{52}=252$, $p<0.001$), a result that agrees with the findings for beginning readers of English. The occurrence of phoneme substitutions on medial vowel segments was, however, less frequent than on initial ($T_{53}=273$, $p<0.001$) or final ($T_{57}=202$, $p<0.001$) consonant segments. Serbo-Croatian differs from English: consonants cause more difficulty for beginning readers than vowels. In an attempt to understand this finding one is reminded that the vowel set in Serbo-Croatian comprises only five vowels and that the Serbo-Croatian vowels are neatly distinctive in the F_1-F_2 plane. On the contrary, within some groups of the Serbo-Croatian consonants the distinctiveness is poor. For example, within the group of four affricates /tʃ/, /tʃj/, /dʒ/, /dʒj/ the phoneme boundaries are extremely fragile. Moreover, in some regions of Yugoslavia the native population replaces the voiced affricates /tʃ/ and /dʒ/ by their respective voiceless mates /tʃj/ and /dʒj/.

In our opinion the result of this experiment indicates that the substitution errors (both the initial consonant and final consonant) were phonetically biased. By far the more frequent errors were the substitutions within the group of the Serbo-Croatian affricates. All proportions of opportunity for substitution in Table 5 are small in comparison with the corresponding figures in the report of Shankweiler and I. Y. Liberman (1972).

A last but not the least interesting finding of this experiment is the fact that the final consonant substitution errors (see Table 5) were more frequent for pseudowords than for words. This suggests that even at an early stage of learning to read the process of decoding is sensitive to lexical content and that the child may possess both nonlexical (orthographic) and lexical routes to the phonology (Baron & Strawson, 1976; Forster & Chambers, 1973; Patterson & Marcel, 1977).

LEXICAL DECISION AND PHONOLOGICAL ANALYSIS

It is commonplace to underscore the fact that English spelling is a less than perfect transcription of the phonology. Nevertheless, English is an alphabet in spite of its apparent phonological capriciousness--for each spelled English word provides strong hints as to its pronunciation. Some students of reading (e.g., Smith, 1971), however, have felt that the hints are so obscure, the relation between script and phonology so opaque, that the fluent reading of English by-passes what must be the complex and arduous process of converting the letter patterns into their related phonological forms. The idea that the fluent reading of English may proceed without reference to the phonology is buttressed by the claim that the English spelling often preserves morphological relatedness, that is, similar meaning (Chomsky, 1970). Given this claim, it is a simple step to supposing that the fluent reading of English proceeds as one might suppose that the fluent reading of logographic writing proceeds, that is, without a phonological intermediary between the printed word and its meaning (e.g., Goodman, 1973).

But forceful arguments can be made and have been made by Rozin and Gleitman (1977) to counter these denials of a phonologic strategy. Indeed, as Rozin and Gleitman (1977) take pains to point out, the observations questioning a phonological mediary cut two ways and when looked at carefully add strength to, rather than weaken, the notion of phonological involvement in the reading of English.

It is evident from what has been said about Serbo-Croatian writing, that neither of the two foregoing arguments against a phonological encoding is especially compelling from the perspective of that orthography. Indeed, if an opaque relation between script and phonology and a preserved transcription of the morphology are advanced as reasons *against* phonological involvement in the reading of English, then a transparent relation between script and phonology and an optimal transcription of the phonology should be received as reasons *for* phonological involvement in the reading of Serbo-Croatian.

At all events, this general issue of the contribution of phonological encoding to reading is given particular expression in various laboratory tasks. An extremely popular task is that of lexical decision, a task in which the subject must decide as rapidly as possible whether a visually presented letter string is a word. A finding often presented as evidence for phonological involvement in accessing English lexical items is that rejection latencies for nonhomophonic pseudowords are shorter than for homophonic pseudowords (Rubenstein, Lewis, & Rubenstein, 1971). That is, it takes longer to initiate response (say, pressing a telegraph key) to indicate "no" (it is not a word) to a pseudoword that sounds exactly like a real word than to a pseudoword that does not sound like any word (also Coltheart, Davelaar, Jonasson, & Besner, 1977). While, in general, lexical decision experiments support the idea of a phonologically mediated access to English lexical items (e.g., Meyer, Schvaneveldt, & Ruddy, 1974), other experiments that use other tasks imply no phonological analysis or, at best, a phonological analysis that occurs subsequent to lexical evaluation (e.g., Green & Shallice, 1976; Kleiman, 1975).

All things considered, however, the emerging orthodoxy appears to be that there is both a phonologically mediated route to the lexicon and a more direct, nonphonological route with the two modes of access relatively independent and possibly parallel in operation. As Gleitman and Rozin (1977) express it, reading probably proceeds at a number of grains of linguistic analysis simultaneously.

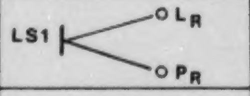
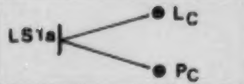
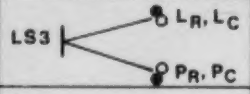
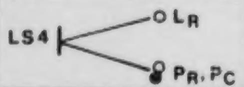
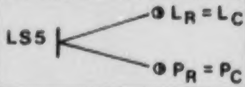
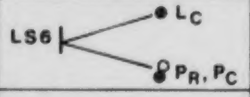
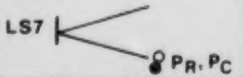
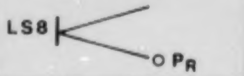
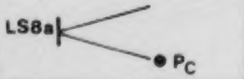
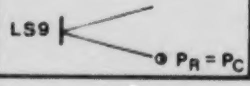
We wish to support the claim of phonological involvement in lexical decision. Evidence is presented that suggests that in lexical decision on Serbo-Croatian letter strings the phonological representation cannot be bypassed and that the phonological interpretation of a letter string is obligatory and automatic. Additionally, evidence is presented to show a complicity between the phonological evaluation and the lexical evaluation of letter strings that is of significance to the construction of a theory of word recognition.

Given the nature of and the relation between the two Serbo-Croatian alphabets it is possible to create a variety of types of letter strings. Thus, a letter string composed of uniquely Roman letters or of uniquely Cyrillic letters (in Figure 1) would receive single phonological interpretation and could be either a word or not a word. In contrast, a letter string composed of the common and ambiguous letters (see Figure 1) would receive two distinct phonological interpretations and could be either a word or not a word; more precisely, it could be a word in one alphabet and a pseudoword in the other or it could represent two different words, one in one alphabet and one in the other.

In a series of three experiments (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978) bi-alphabetic subjects were invited--by experimental design and by instruction--to relate to letter strings (block capitals) in the Roman alphabet mode. None of the letter strings seen by a subject were comprised of uniquely Cyrillic letters and relatively few of the letter strings were composed of common and ambiguous letters, that is to say, could even be read as Cyrillic. The conclusion on which all three experiments converged was that lexical decision to a letter string was slower when that string could be given two phonological readings (that is, could be read in either the assigned Roman alphabet mode or the nonassigned Cyrillic alphabet mode) but if and only if the letter string was a word in at least one of the alphabets. Pseudowords that could be read in both alphabets were rejected no slower than pseudowords constructed from the set of letters unique to the Roman alphabet.

This result is nicely illustrated by a recent experiment in which there was no imposed alphabet bias: The adult bi-alphabetic subject (there were 48 subjects in the experiment) decides whether a string of (capital) letters is a word in the Serbo-Croatian language. In this experiment, unlike the previous ones, letter strings containing uniquely Roman letters and letter strings containing uniquely Cyrillic letters were presented. The types of letter strings (LS) examined are shown in Table 6 together with the correct lexical decision for each type. (The odd labeling of letter strings is to maintain consistency with the table of letter strings given previously in Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978; the present table is more inclusive). Table 6 is self-explanatory although it needs remarking that LS5

Table 6. Types of letter strings in the Roman and Cyrillic alphabet.

Type of letter string (LS)	Lexical entry (L)		Phonological representation (P)		Symbolic representation	Is it a word? (in Roman or in Cyrillic)
	In Roman (L_R)?	In Cyrillic (L_C)?	In Roman (P_R)?	In Cyrillic (P_C)?		
LS1	Yes	No	Yes	No		Yes
LS1a	No	Yes	No	Yes		Yes
LS3	Yes	Yes	Yes	Yes	 LS3 L_R, L_C P_R, P_C	Yes
LS4	Yes	No	Yes	Yes		Yes
LS5	Yes	Yes	Yes	Yes	 LS5 $L_R = L_C$ $P_R = P_C$	Yes
LS6	No	Yes	Yes	Yes		Yes
LS7	No	No	Yes	Yes		No
LS8	No	No	Yes	No		No
LS8a	No	No	No	Yes		No
LS9	No	No	Yes	Yes		No

and LS9 are composed solely from the common letters (see Figure 1) and are therefore read the same way and mean the same thing (in the case of LS5) in Roman and Cyrillic. The results of the experiment are shown in Figure 2. It is apparent from inspection of Figure 2 that lexical decision was impaired for those letter strings that could be given both Cyrillic and Roman interpretations but only if the letter string was a word. To give two of the relevant comparisons, decision times to LS4 were significantly slower than decision times to LS1 (F=11.72; df=1,26; $p<0.01$); decision times to LS3 were significantly slower than decision times to LS2 (F=33.4; df=1,27; $p<0.001$). The latter contrast is especially interesting since letter strings of type LS3 are words in both alphabets and since a general observation in the literature on English words is that letter strings with multiple meanings are accepted as words faster than letter strings with a single meaning (e.g., Jastrzembski & Stanners, 1975). Clearly, the present observation is counter to this general finding. It should also be noted that the slower decision time to LS3 was witnessed in our previous research (Lukatela, Savić, Gligorijević, Ognjenović & Turvey, 1978). Returning to the data represented by Figure 2, where the letter string was not a word, the lexical decision was not retarded by phonological bivalence: decision times to LS7 did not differ, for example, from those to LS8(F=2.44, df=1,50).

As anticipated, these data on bi-alphabetic lexical decision permit two conclusions of some significance to an understanding of the reading of Serbo-Croatian. (We are assuming like others--for example, Coltheart et al., 1977--that lexical decision is a laboratory task well suited to investigating the nature of the information extracted from a printed word for use of lexical access.) First, the data suggest strongly that phonological encoding of Serbo-Croatian words is an automatic and extremely rapid process; as we have seen, phonological bivalence interferes with lexical decision. Second, the data suggest that it is not phonological bivalence per se that retards lexical decision, rather the necessary contingency is that the phonologically bivalent letter string being evaluated must be a word in the Serbo-Croatian language.[1]

There are a number of theories that could be pursued by way of explaining this curious result of bi-alphabetic lexical decision. They are not pursued here for there is little to be gained at this stage by adjusting the details of this or that account of lexical decision (e.g., Coltheart et al., 1977; Meyer & Ruddy, Note 1) so as to force a fit with the present data. It suffices, perhaps, to note the Coltheart et al. (1977) concluding lament that for English there is no compelling evidence for the view that the mapping from printed word to lexical entry references the phonology. They propose that:

> Unequivocal evidence for this view would be obtained by demonstrating that the phonological code for a word is sometimes used in making the "yes" response to that word in a lexical decision or categorization task; such a demonstration remains to be achieved (Coltheart et al., 1977, p. 551).

Do the present data constitute such a demonstration for Serbo-Croatian?

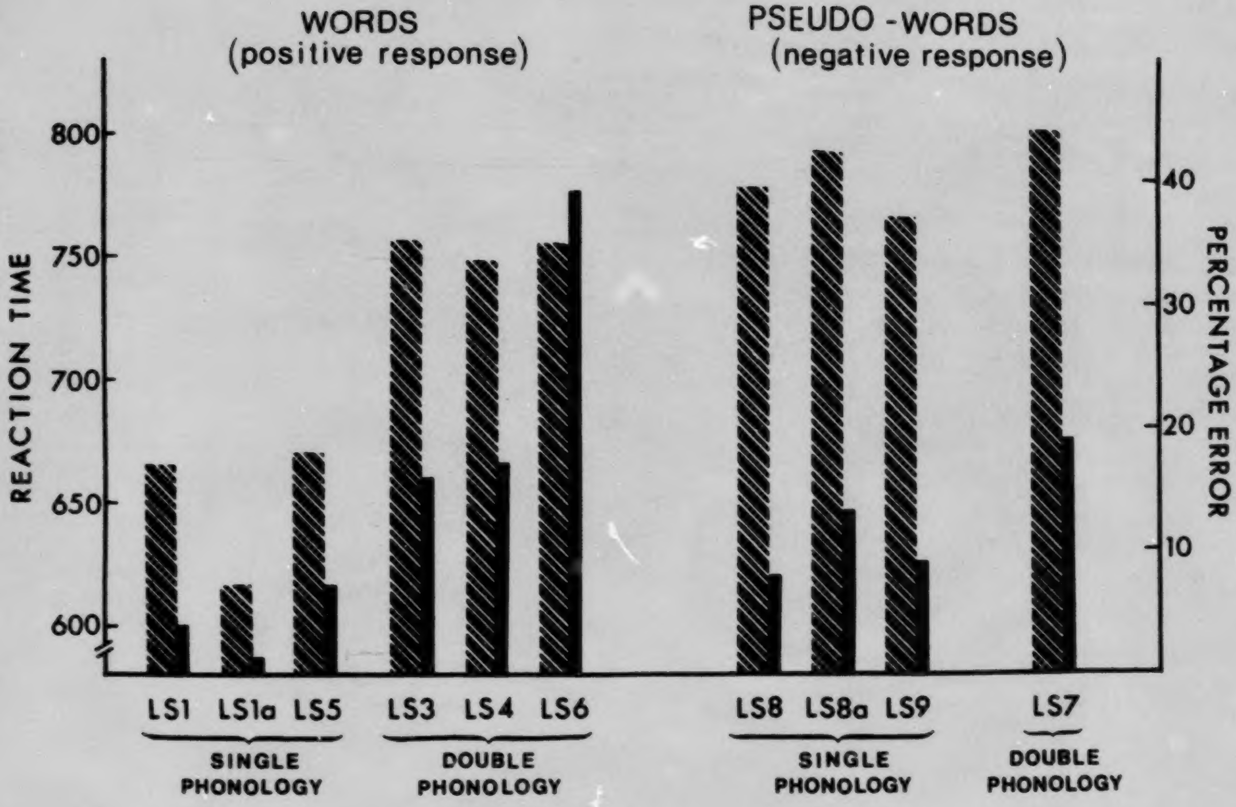


Figure 2. Lexical decision latencies and errors for Serbo-Croatian letter strings that are readable in only one alphabet or readable in both alphabets.

THE PROCESSING RELATION BETWEEN THE TWO SERBO-CROATIAN ALPHABETS

A question that has been pursued at some length is how the Roman and Cyrillic alphabets relate psychologically. For the reader of Serbo-Croatian the alphabets must be kept distinct at some level (or in some manner) of processing in order to circumvent the ambiguous characters as a potential source of phonetic confusion. Might we therefore speak of an alphabet mode implying perhaps that the reader can be in one mode or the other but not in both concurrently? The experiments just described bear on this question.

And how are the two alphabets memorially represented? If there are two alphabet spaces are all the letters of the Roman alphabet stored in one space and all the letters of the Cyrillic alphabet stored in the other? Or is there a region of overlap, say, the representations of the common letters? Given that the meaning of one alphabet precedes the other, how is priority in learning manifest in either the processing or the representation of the two alphabets? These questions and others guided our attempts to understand the psychological fit between the two Serbo-Croatian writing systems (Lukatela, Savić, Ognjenović, & Turvey, 1978); a part of that research is reported here.

A very simple experiment proved exceptionally instructive. Native Eastern Yugoslavians (those who learn Cyrillic first) were presented individual Roman and Cyrillic letters in random order and pressed a key as quickly as possible in answer to the question "Is this letter Cyrillic?" or to the question "Is this letter Roman?" The results are given in Figure 3. It took considerably longer to verify the common letters (see Figure 1) were Roman in the "Is this letter Roman?" condition than to verify that the common letters were Cyrillic in the "Is this letter Cyrillic?" condition. The suggestion is that the subjects of the experiment viewed the common letters as essentially members of the Cyrillic alphabet and only indirectly as members of the Roman alphabet. Arguing in like style, the ambiguous characters would appear to inhabit both alphabet spaces. The most telling observation however was this: rejecting Cyrillic letters in the Roman alphabet mode took appreciably longer than rejecting Roman letters in the Cyrillic alphabet mode.

We have come to look at these data in the following way. We reasoned that the average latency for rejecting a Cyrillic character as Roman is an index of the degree to which a description of a Cyrillic character is, on the average, similar to a description of a Roman character. In the notation of Tversky (1977) this similarity may be written as $s(c,r)$ where the perceptual representation of the target Cyrillic letter (c) is the *subject* of the relation and where the memorial representation of an individual Roman letter (r) is the *referent*. Similarly, the average latency for rejecting a Roman character as Cyrillic indexes $s(r,c)$. It follows, therefore, that $s(c,r)>s(r,c)$. In other words, for speakers of Serbo-Croatian who have learned the Cyrillic alphabet first, the perceptual descriptions of Cyrillic characters are, on the average, more similar to the memorial descriptions of Roman characters than the perceptul descriptions of Roman characters are, on the average, similar to the memorial descriptions of Cyrillic characters.

What is the basis for this asymmetry? By Tversky's (1977) argument asymmetric similarities such as X is more similar to Y than vice versa hold if and only if Y, the referent term, is more salient on some nontrivial dimension

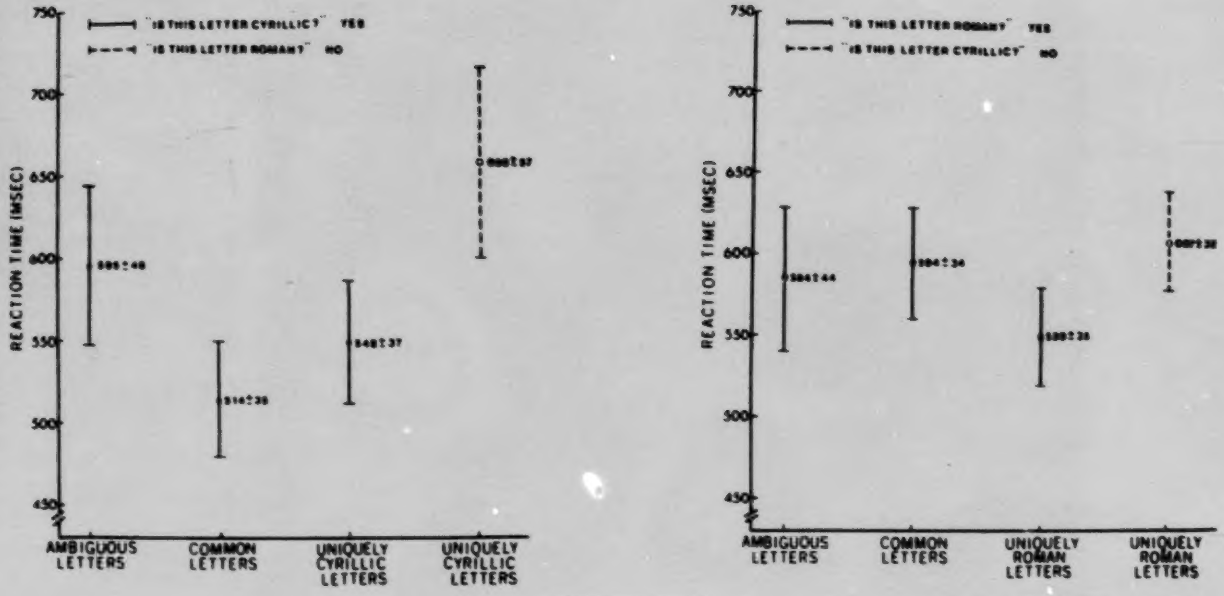


Figure 3. Mean latencies and their range of variation for the alphabet decision task performed by subjects who learned the Cyrillic alphabet first.

from X, the subject term. The putative salience of (processing) the Roman alphabet may arise because the dimensions of description of the Roman alphabet include those of the Cyrillic; or that the descriptors of the Roman alphabet distinguish the Roman characters more efficiently than the descriptors of the Cyrillic alphabet distinguish Cyrillic characters. In short, the basis for the asymmetry may lie in some absolute property distinguishing the structure of the two alphabets. If true, the direction of the asymmetry should be indifferent to the order in which the alphabets are acquired. On the other hand, the basis for the asymmetry may just be the order of acquisition. To this purpose, the alphabet-decision task described above was replicated with subjects who had acquired the Roman alphabet first and the Cyrillic alphabet second. The results are shown in Figure 4. They reveal that under the two question regimes ("Is this letter Roman?"; "Is this letter Cyrillic?") these subjects behaved differently, as did the subjects in the first experiment. But most importantly the behavior of the subjects indigenous to Western Yugoslavia was diametrically opposite to that of the subjects indigenous of Eastern Yugoslavia (compare Figure 4 with Figure 3). By the same reasoning as outlined above we conclude, for subjects who learned the Roman alphabet first, that $s(r,c)>s(c,r)$. That is, for Roman-first subjects, processing Roman letters is more similar to processing Cyrillic letters than vice versa. More generally we conclude that the alphabet-processing asymmetry is owing not to a fixed structural property of the alphabets but to their order of acquisition. One tentative conclusion to be drawn is that the procedure developed by the child to decode the letters of the first acquired alphabet is modified for the second acquired alphabet so that decoding the second acquired alphabet necessarily entails the procedure for decoding the first acquired alphabet but not vice versa.

But perhaps the more outstanding, although equally tentative, conclusion to be drawn is that the order in which the alphabets are acquired, and the concomitant early bias in reading toward one of the alphabets, leaves a profound impression on the letter decoding processes of adult readers of Serbo-Croatian. This conclusion is not unrelated to some results recently published by Jackson and McClelland (1979). In the view of some students of reading (e.g. Kolers, 1969; Smith, 1971) individual differences in the reading ability of experienced readers are solely differences in comprehension ability. The research of Jackson and McClelland brings this view into question by showing individual differences in the ability of American college student readers to access letter codes, an ability that accounts for a significant portion of the variance in effective reading speed. What has been noted with mature Serbo-Croatian readers is that in the alphabet decision task there is an interaction between the alphabet first learned and the alphabet being decided upon. The pattern of decision times for Roman-first subjects is, on the significant contrasts, a mirror image of the pattern for the Cyrillic-first subjects. What is surprising about this interaction is that the subjects have been reading in the two alphabets for between 12 and 16 years and yet on a simple decision task the alphabet learned first makes its mark. The point on which our data and those of Jackson and McClelland would appear to converge is that the basic encoding processes by which letters of the alphabet are distinguished and named are not necessarily asymptotic in mature readers; nor is mature reading indifferent, perhaps, to the manner of their acquisition.

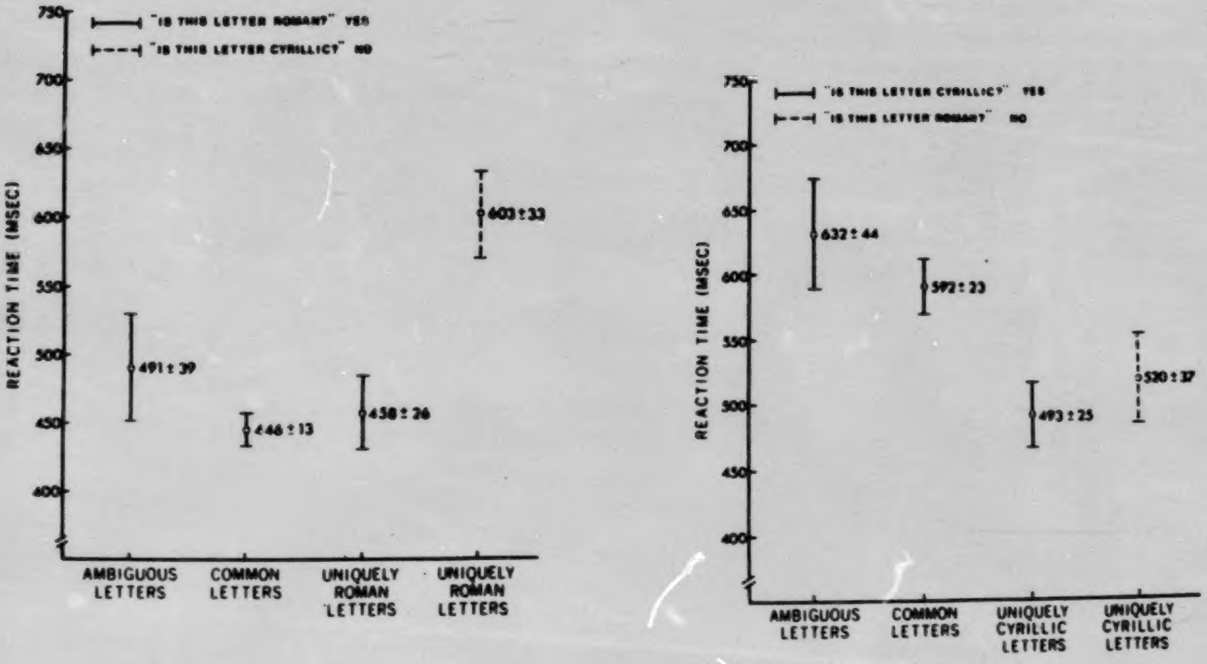


Figure 4. Mean latencies and their range of variation for the alphabet decision task performed by subjects who learned the Roman alphabet first.

REFERENCE NOTE

1. Meyer, D. E., & Ruddy, M. G. *Lexical-memory retrieval based on graphemic and phonemic representations of printed words*. Paper presented at the meetings of the Psychonomic Society, St. Louis, 1973.

REFERENCES

Baron, J., & Strawson, C. Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 386-393.

Carroll, J. B. Defining language comprehension: Some speculations. In J. B. Carroll & R. O. Freedle (Eds.), *Language comprehension and the acquisition of knowledge*. Washington, D.C.: V. H. Winston, 1972.

Chomsky, N. Phonology and reading. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books, 1970.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI*. London: Academic Press, 1977.

Elkonin, D. B. The psychology of mastering the elements of reading. In B. Simon & J. Simon (Eds.), *Educational psychology in the U.S.S.R.* London: Routledge & Kegan Paul, 1963.

Elkonin, D. B. U.S.S.R. In J. Downing (Ed.), *Comparative reading*. New York: Macmillan, 1973.

Forster, K. K., & Chambers, S. M. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 627-635.

Francis, W. N. Linguistics and reading. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books, 1970.

Gibson, E. J., & Levin, H. *The psychology of reading*. Cambridge: The MIT Press, 1975.

Gibson, E. J., Pick, A., Osser, H., & Hammond, M. The role of grapheme-phoneme correspondence in the perception of words. *American Journal of Psychology*, 1962, *75*, 554-570.

Gleitman, L. R., & Rozin, P. The structure and acquisition of reading I: Relations between orthographies and the structure of language. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY conference*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Goodman, K. S. The 13th easy way to make learning to read difficult: A reaction to Gleitman and Rozin. *Reading Research Quarterly*, 1973, *8*, 484-493.

Gray, W. S. *The teaching of reading and writing, an international survey*. Paris: UNESCO, 1956.

Green, D. W., & Shallice, T. Direct visual access in reading for meaning. *Memory & Cognition*, 1976, *4*, 753-758.

Henderson, L. Word recognition. In N. S. Sutherland (Ed.), *Tutorial essays in experimental psychology*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Jackson, M. D., & McClelland, J. L. Processing determinants of reading speed. *Journal of Experimental Psychology: General*, 1979, *108*, 151-181.

Jastrzembski, J. E., & Stanners, R. Multiple-word meanings and lexical search speed. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 534-537.

Karadžić, V. S. Pismenica serbskoga jezika po govoru prostoga naroda. Wien: J. Schnirer Press, 1814.

Kleiman, G. M. Speech recoding in reading. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 323-339.

Kolers, P. A. Reading is only incidentally visual. In K. S. Goodman & J. Fleming (Eds.), *Psycholinguistics and the teaching of reading*. Newark, Del.: International Reading Association, 1969.

Liberman, I. Y., & Shankweiler, D. Speech, the alphabet, and teaching to read. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and practice of early reading* (Vol. 2). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY conference*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. Bi-alphabetical lexical decision. *Language and Speech*, 1978, *21*, 142-165.

Lukatela, G., Savić, M., Ognjenović, P., & Turvey, M. T. On the relation between processing the Roman and Cyrillic alphabets: A preliminary analysis with bi-alphabetical readers. *Language and Speech*, 1978, *21*, 113-141.

Mattingly, I. G. Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge: The MIT Press, 1972.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Loci of contextual effects on visual word recognition. In S. Dornic & P. M. A. Rabbitt (Eds.), *Attention and performance V*. London: Academic Press, 1974.

Patterson, K. E., & Marcel, A. J. Aphasia, dyslexia and the phonological coding of written words. *Quarterly Journal of Experimental Psychology*, 1977, *29*, 307-318.

Rosner, J., & Simon, D. P. The auditory analysis test: An initial report. *Journal of Learning Disorders*, 1971, *4*, 40-48.

Rozin, P., & Gleitman, L. R. The structure and acquisition of reading II: The reading process and the acquisition of the alphabetic principle. In A. S. Reber & D. L. Scarborough (Eds.), *Towards a psychology of reading: The proceedings of the CUNY conference*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. Evidence for phonetic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 647-657.

Savin, H. B. What the child knows about speech when he starts to learn to read. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationship between speech and reading*. Cambridge, Mass.: The MIT Press, 1972.

Shankweiler, D., & Liberman, I. Y. Misreading: A search for causes. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationship between speech and reading*. Cambridge, Mass.: The MIT

Press, 1972, 293-317.

Smith, F. Understanding reading: A psycholinguistic analysis of reading and learning to read. New York: Holt, 1971.

Tversky, A. Features of similarity. Psychological Review, 1977, 84, 327-352.

Venezky, R. L. The structure of English orthography. The Hague: Mouton, 1970.

Venezky, R. L. The letter-sound generalizations of first, second, and third grade Finnish children. Journal of Educational Psychology, 1973, 64, 288-292.

Venezky, R. L., & Massaro, D. W. The role of orthographic regularity in word recognition. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1). Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. Psychological Review, 1970, 77, 1-15.

FOOTNOTES

[1] In a subsequent analysis of these data (see Lukatela, Popadić, Ognjenović, & Turvey, this volume), the detriment to performance incurred by phonologically bivalent letter strings occurred both for words and pseudowords.

LEXICAL DECISION IN A PHONOLOGICALLY SHALLOW ORTHOGRAPHY[*]

G. Lukatela[+], D. Popadić[+], P. Ognjenović[+], and M. T. Turvey[++]

Abstract. The Serbo-Croatian language is written in two alphabets, Roman and Cyrillic. Both orthographies transcribe the sounds of the language in a regular and straightforward fashion and may, therefore, be referred to as phonologically shallow in contrast to English orthography, which is phonologically deep. Most of the alphabet characters are unique to one alphabet or the other. There are, however, a number of shared characters, some of which receive the same reading and some of which receive a different reading, in the two alphabets. It is possible, therefore, to construct a variety of types of letter strings. Some of these can be read in only one way and can be either a word or nonsense. Other letter strings can be pronounced one way if read as Roman and in a distinctively different way if read as Cyrillic and can be words in both alphabets--but different words; or they can be nonsense in both alphabets or nonsense in one alphabet and a word in the other. In a lexical decision task conducted with bialphabetical readers, it was shown that words that can be read in two different ways are accepted more slowly and with greater error than words that can be read only one way. It was concluded that for the phonologically shallow writing systems of Serbo-Croatian, lexical decision proceeds with reference to the phonology.

A case can be made for distinguishing among alphabetic writing systems in terms of the derivational complexity that relates the spelling to the underlying phonological form (Liberman, Liberman, Mattingly, & Shankweiler, 1980). English orthography is the notorious example of a "phonologically deep" writing system; but it is a truly phonographic orthography in spite of its depth because each spelled English word contains strong hints as to its pronunciation. Nevertheless, the opaqueness of the link between English script and phonology is seen by many as a barrier to phonological involvement in fluent reading (Goodman, 1973; Kolers, 1970; Smith, 1971). The argument runs as follows: Given the difficulty of deriving the phonology, readers of English would be considerably better off if they had the option of bypassing the phonology and of relating to their alphabetic orthography much in the same way that the readers of Chinese, say, are thought to relate to their logographic orthography, that is, of proceeding directly from script to

---

[*] Also in *Memory & Cognition*, 1980, *8*, 124-132.

[+] University of Belgrade, Belgrade, Yugoslavia.

[++] Also University of Connecticut, Storrs.

Acknowledgment: This research was supported in part by NICHD Grant HD-08495 to the University of Belgrade and in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

meaning. The latter point of view receives some measure of support from analyses that purportedly reveal a closer fit of English orthography to morphology rather than to phonology (e.g., Chomsky, 1970).

The generally voiced arguments for denying a phonological intermediary in the fluent reading of English have been carefully reviewed by Rozin and Gleitman (1977). Their impression is that these arguments cut both ways and can, ironically, be taken to strengthen rather than to weaken the claim for a principled use of phonology in reading. Additionally, Rozin and Gleitman (1977) point out that it is wiser to interpret the English writing system as a rich mixture of several grains of linguistic representation peppered with arbitrary features (arising from scribal practices, printers' conventions, etc.) rather than as a spelling system that is optimal for any single grain of linguistic representation.

One implication of the last remark is that the reading of English may proceed simultaneously at several grain sizes of linguistic analysis (Rozin & Gleitman, 1977). It is, therefore, easy to venture that the multiple linguistic analyses afforded by English writing are reason enough for the failure to achieve experimental resolution to the question of a phonological mediary in the mapping from script to meaning. In any given experimental situation, the phonological representation may be obscured by other permissible representations. On the other hand, or additionally, it can be ventured that the failure to resolve the question of phonological mediation is owing to the fact that most of the experimental procedures used to investigate it are not directly relevant to its resolution. Coltheart and his colleagues (Coltheart, Davelaar, Jonasson, & Besner, 1977; Davelaar, Coltheart, Besner, & Jonasson, 1978) have argued that the only legitimate experimental tasks are those that logically require the use of lexical knowledge. The lexical decision task meets the advocated criterion: Letter strings that are words must be rapidly distinguished from letter strings that are pseudowords.

One consistent finding from lexical decision research that is interpreted by some as implicating phonological involvement in the accessing of English lexical items is that it takes an adult reader longer to reject a pseudoword that sounds exactly like a real word than to reject a pseudoword that does not sound like any word (Coltheart et al., 1977; Rubenstein, Lewis, & Rubenstein, 1971). Importantly, however, a cognate observation has proven less reliable, namely, that acceptance latencies are slower for homophonous words than for nonhomophonous words (Rubenstein et al., 1971). When differences in parts of speech and frequency of occurrence are ruled out, words that sound like other words are accepted as rapidly as words that are phonetically dissimilar to other words (Coltheart et al., 1977). In summary, it would appear that phonology mediates the rejection of pseudowords but does not mediate the acceptance of words, a conclusion that undercuts the claim that phonology mediates the normal reading of English. In paraphrase of Coltheart et al. (1977), evidence for phonologically mediated lexical access would be more convincing if phonological involvement could be shown in positive lexical decisions.

Although the sought-after evidence has been forthcoming, it has not been without an important qualification. Davelaar et al. (1978) demonstrated that homophony affected lexical decision on words but only when the pseudowords, the distractor items, if you wish, were nonhomophonic with lexical items. We see, in short, that phonological involvement in the accessing of English lexical items may well be optional. Apparently, when the strategy of referencing the phonology is less than ideal, as in the case in a lexical decision task in which the pseudowords sound like real words, the strategy can be inhibited and other strategies, other grains of linguistic analysis, are given prominence (cf. Davelaar et al., 1978).

The focus of the present paper is a language that is written in a "phonologically shallow" orthography. Serbo-Croatian, the major language of Yugoslavia, is written in two alphabets, Roman and Cyrillic, both of which were constructed in the last century according to the simple rule: "Write as you speak and speak as it is written." Both the Roman and Cyrillic orthographies transcribe the sounds of the Serbo-Croatian language in a regular and straightforward fashion, and there are no (nontrivial) derivation rules to speak of. (Indeed, it is questionable whether the notion of "phonological representation" is befitting the written Serbo-Croatian language. "Phonetic representation" may be sufficient, and more suitable.) (1)

It seems to us that the generally expressed reasons given against a phonological mediary in the fluent reading of English are not applicable, even in principle, to the fluent reading of Serbo-Croatian (Lukatela & Turvey, 1980). The Serbo-Croatian orthographies are optimal for transcribing the phonology and are transparent in that regard; therefore, no special difficulty is raised for a phonological mediary in the reading of Serbo-Croatian. We might suppose, therefore, that lexical decision on Serbo-Croatian letter strings exhibits a greater or, at least, a more apparent sensitivity to phonology than does lexical decision on English letter strings. Previous research with Serbo-Croatian (Lukatela, Savić, Gligorijerić, Ognjenović, & Turvey, 1978) might be interpreted as evidence of an obligatory phonological reference in lexical decision, but we must, of necessity, preface a summary of that research by a brief statement of the relation between the two Serbo-Croatian alphabets. (For a more detailed description, see Lukatela, Savić, Ognjenović, & Turvey, 1978; Lukatela & Turvey, 1980).

The Roman and Cyrillic alphabets map onto the same set of phones but comprise two sets of letters that are, with certain exceptions, mutually exclusive (see Figure 1). Most of the Roman and Cyrillic letters are unique to their respective alphabets. There are, however, a number of letters that the two alphabets have in common. The phonemic interpretation of some of these shared letters is the same whether they are read as Cyrillic or as Roman letters; these are referred to as common letters. Other members of the shared letters have two phonemic interpretations, one in the Roman reading and one in the Cyrillic reading; these are referred to as ambiguous letters. Whatever their category the individual letters of the two alphabets have phonemic interpretations that are virtually invariant over letter contexts. Moreover, all the individual letters in a string of letters, be it a word or nonsense, are pronounced--there are no letters made silent by context. Finally, but not least in importance, we should note that the two alphabets are used competently by a large portion of the population. This is due, in part, to an educational requirement that both alphabets be taught within the first two grades. The first-taught alphabet is Roman in the western part of Yugoslavia and Cyrillic in the eastern part of Yugoslavia.

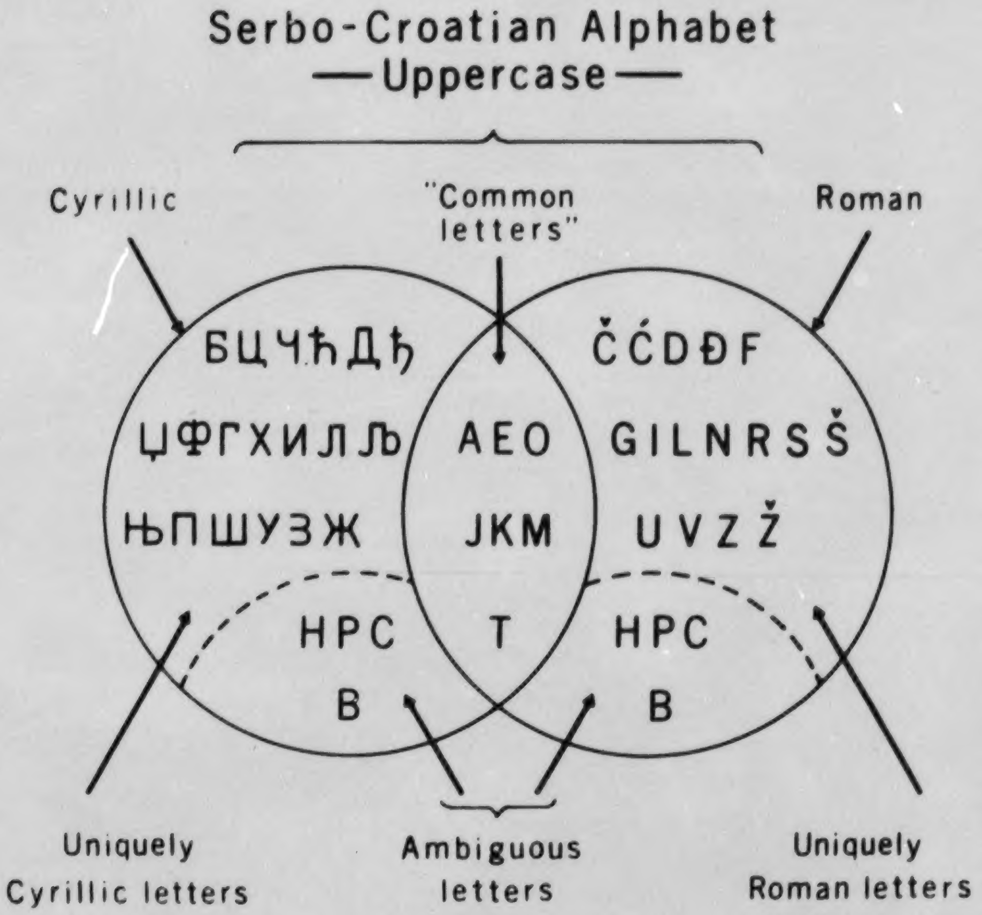

Serbo-Croatian Alphabet
—Uppercase—
Cyrillic
"Common letters"
Roman
БЦЧЋДЂ
ЏФГХИЛЉ
ЊПШУЗЖ
AEO
JKM
T
ČĆDĐF
GILNRSŠ
UVZŽ
HPC
B
HPC
B
Uniquely Cyrillic letters
Ambiguous letters
Uniquely Roman letters

Figure 1. The uppercase characters of the Roman and Cyrillic alphabets of Serbo-Croatian.

Given the nature of and the relation between the two Serbo-Croatian alphabets, it is possible to construct a variety of types of letter strings. A letter string of uniquely Roman letters or of uniquely Cyrillic letters would be read in only one way and could be either a word or nonsense. A letter string composed of the common and ambiguous letters could be pronounced one way if read as Roman and pronounced in a distinctively different way if read as Cyrillic; moreover, it could be a word in one alphabet and nonsense in the other, or it could represent two different words, one in one alphabet and one in the other, or it could be nonsense in both alphabets.

We can now summarize our previous research on lexical decision. In three experiments, subjects who could read in both alphabets and who had received their elementary education in eastern Yugoslavia were presented letter strings for lexical decision in the Roman alphabet mode. The requisite mode was determined by instruction and by the selection of letter strings. Letters unique to the Cyrillic alphabet were not used to compose the letter strings and comparatively few of the letter strings were constructed from the common and ambiguous letters. In short, very few of the presented letter strings could be read in the Cyrillic alphabet mode. It was demonstrated that lexical decision was slowed when a letter string could be read in two ways (i.e., could be read in either the assigned Roman alphabet or the nonassigned Cyrillic alphabet), but only if it were the case that the letter string was in fact a word in (at least) one of the alphabets. A nonsense string of letters readable in both alphabets was rejected no more slowly than a nonsense string constructed from the set of letters unique to the Roman alphabet.

By arranging matters so as to make the use of a phonological code punitive in accessing English lexical items, Davelaar et al. (1978) found that phonological access was abandoned or that, if it was used, its consequences were ignored. In the Lukatela, Savić, Gligorijević, Ognjenović, and Turvey (1978) experiments, matters were arranged so that only one phonological code, that related to the Roman alphabet, was necessary for the successful performance of the task. But our subjects, apparently, were unable to suppress the alternative (and uncalled for) phonological code, that related to the Cyrillic alphabet.

That a familiar item may be encoded automatically, in the related senses of not requiring conscious attention and of not being optional, is central to certain contemporary views of attention and pattern recognition, of which that of Posner and Snyder (1975) is a notable example.

In the experiment reported in the present paper, bialphabetical subjects made lexical decisions on letter strings that were composed from the unique letters of both alphabets as well as from the common and ambiguous letters. That is to say, in contrast with the previous experiments (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978) no alphabet bias was imposed upon the subjects by the selection of letter strings; nor was it imposed by instruction. Subjects simply had to identify whether or not a letter string, be it Cyrillic or Roman, represented a word in the Serbo-Croatian language. On the evidence of our previous research, it would be nonoptimal to access the lexicon via the phonology if that means of access necessarily entailed both the Roman and the Cyrillic phonological codes. Far more prudent would be a strategy in which access to the lexicon was restricted to the graphemic route (see Coltheart et al., 1977; Meyer, Schavaneveldt, & Ruddy, 1974) or, at least, a strategy in which, of the two routes, only the graphemic was heeded in final decision making. It proves to be the case, however, that, consonant with the earlier observations on biased bialphabetical subjects, unbiased bialphabetical subjects, under the conditions of the present experiment, exhibit an inability to suppress the phonological coding of Serbo-Croatian letter strings. As before, words that can be read in two different ways are accepted more slowly and with greater error than words that can be read only one way.

METHOD

Subjects

The participants in the experiments were 48 students from the Department of Psychology at the University of Belgrade. The majority of the 48 students had received their elementary education in eastern Yugoslavia, and all of them had participated previously in reaction time experiments.

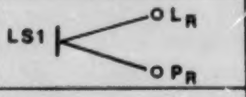
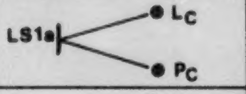
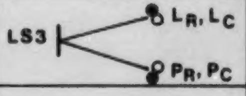
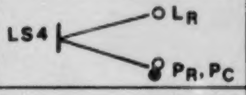
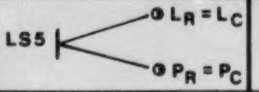
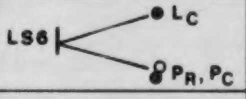
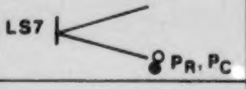
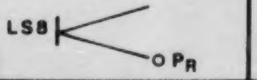
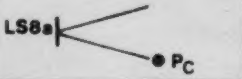
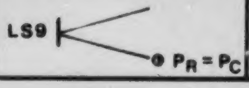
Materials and Design

Letraset black uppercase Roman and Cyrillic letters (Helvetia Light, 12 point) were used to prepare the letter strings. A string of three to six letters arranged horizontally at the center of a 35-mm slide represented a word or a pseudoword in the Serbo-Croatian language. There are no frequency counts for Serbo-Croatian words comparable to the Thorndike-Lorge or Kučera-Francis counts for English words. As with our previous experiments, all words were selected from the middle range of word frequencies for Serbian elementary school children, as reported by Lukić (1970). The words readable in only one alphabet were chosen so that their mean frequencies of occurrence were as close as possible to those of the words readable in both alphabets. While it is possible that words selected from the Lukić table of frequencies may not be either as close together or as far apart on a table of frequencies of adult usage, it is most unlikely that, where differences in frequency arise, those differences are in terms of the single-alphabet/double-alphabet distinction. The point we wish to underscore is that there is little reason to believe that in adult usage the bialphabetic words of the present experiment occur less frequently than the single-alphabet words of the present experiment.

In addition to the frequency constraint, word selection was restricted to words that did not contain rare consonant clusters. That restriction was also applied to the pseudoword letter strings that were the same length and the same number of syllables as the words. All in all, there were 10 different types of letter strings (LS); these are shown in Table 1, together with the correct lexical decision for each type. (The reason for the odd labeling of the letter strings is to maintain consistency with the table of letter strings given previously in Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978; the present table includes letter strings that are uniquely Cyrillic, which the previous table did not.) Table 1 is largely self-explanatory, but one useful point of clarification is that LS5 and LS9 are constructed solely from the common letters (see Figure 1) and are therefore read the same way and, if

Table 1

Type of letter strings in the Roman and Cyrillic Alphabets

Type of letter string (LS)	Lexical entry (L)		Phonological representation (P)		Symbolic representation	Is it a word? (in Roman or in Cyrillic)
	In Roman (L_R)?	In Cyrillic (L_C)?	In Roman (P_R)?	In Cyrillic (P_C)?		
LS1	Yes	No	Yes	No	 LS1 ○ L_R ○ P_R	Yes
LS1a	No	Yes	No	Yes	 LS1a ● L_C ● P_C	Yes
LS3	Yes	Yes	Yes	Yes	 LS3 ○ L_R, L_C ○ P_R, P_C	Yes
LS4	Yes	No	Yes	Yes	 LS4 ○ L_R ○ P_R, P_C	Yes
LS5	Yes	Yes	Yes	Yes	 LS5 ◐ $L_R = L_C$ ◐ $P_R = P_C$	Yes
LS6	No	Yes	Yes	Yes	 LS6 ● L_C ○ P_R, P_C	Yes
LS7	No	No	Yes	Yes	 LS7 ○ P_R, P_C	No
LS8	No	No	Yes	No	 LS8 ○ P_R	No
LS8a	No	No	No	Yes	 LS8a ● P_C	No
LS9	No	No	Yes	Yes	 LS9 ◐ $P_R = P_C$	No

Note--Read open circles as Roman interpretation and closed circles as Cyrillic interpretation.

words, mean the same thing in the Roman and in the Cyrillic alphabets. In total, 144 letter strings were constructed, of which half were words (12 tokens for each of the six types of word letter string) and half were pseudowords (18 tokens for each of the four types of pseudoword letter string).

The 144 letter strings seen by a subject were presented in four blocks. In each block the letter strings of each type were presented in a pseudorandom order. The sequence of blocks was balanced across subjects, and the same string of letters was never judged more than once by a subject.

Procedure

The subject was seated at a three-channel tachistoscope (Scientific Prototype, Model GB). The subject was instructed to focus on the fixation point in the center of a preexposure field that was present at all times except during presentation of a letter string. Each letter string was preceded by an auditory warning signal. The onset of a letter string triggered an electronic counter that was stopped when the subject pressed one of two buttons on a response panel in front of him. Both hands were used. Both thumbs were placed on a telegraph key close to the subject, and both forefingers were placed on another telegraph key 2 in. further away. The subject depressed the closer key (thumbs) if the letter string was a pseudoword and the other further key (forefingers) if the letter string was a word. Regardless of the subject's response time, a letter string was always automatically replaced after 750 msec by the preexposure field.

RESULTS

The decision latency of each subject to each type of letter string was the basic datum for analysis. Those responses that exceeded 1,300 msec were considered errors ("slow responses"), together with "regular" errors, namely, those responses in which the wrong decision was made. A lower criterion of 250 msec was also applied to rule out excessively fast responses, but no responses of this rapidity occurred in the experiment. For purposes of analysis, the latency of a subject's incorrect response was replaced by his or her average latency for that particular type of letter string. Figure 2 gives the decision time and error data for the 10 types of letter strings. The analysis of variance conducted on the data included three factors: The type of letter string was treated as a fixed factor, with words and subjects treated as random factors. The relevant comparisons follow.

First, we consider the analysis of positive decision times. Decision latency was significantly slower (1) for letter strings of Type LS4 than for letter strings of Type LS1 [$F(1,26)=11.72$, $p<.01$], (2) for letter strings of Type LS6 than for letter strings of Type LS1a [$F(1,25)=41.55$, $p<.001$], (3) for letter strings of Type LS3 than for letter strings of Type LS5 [$F(1,27)=8.90$, $p<.01$].

With regard to the total errors (both slow and regular) on positive response trials, a Wilcoxon signed-ranks test was conducted on the proportions

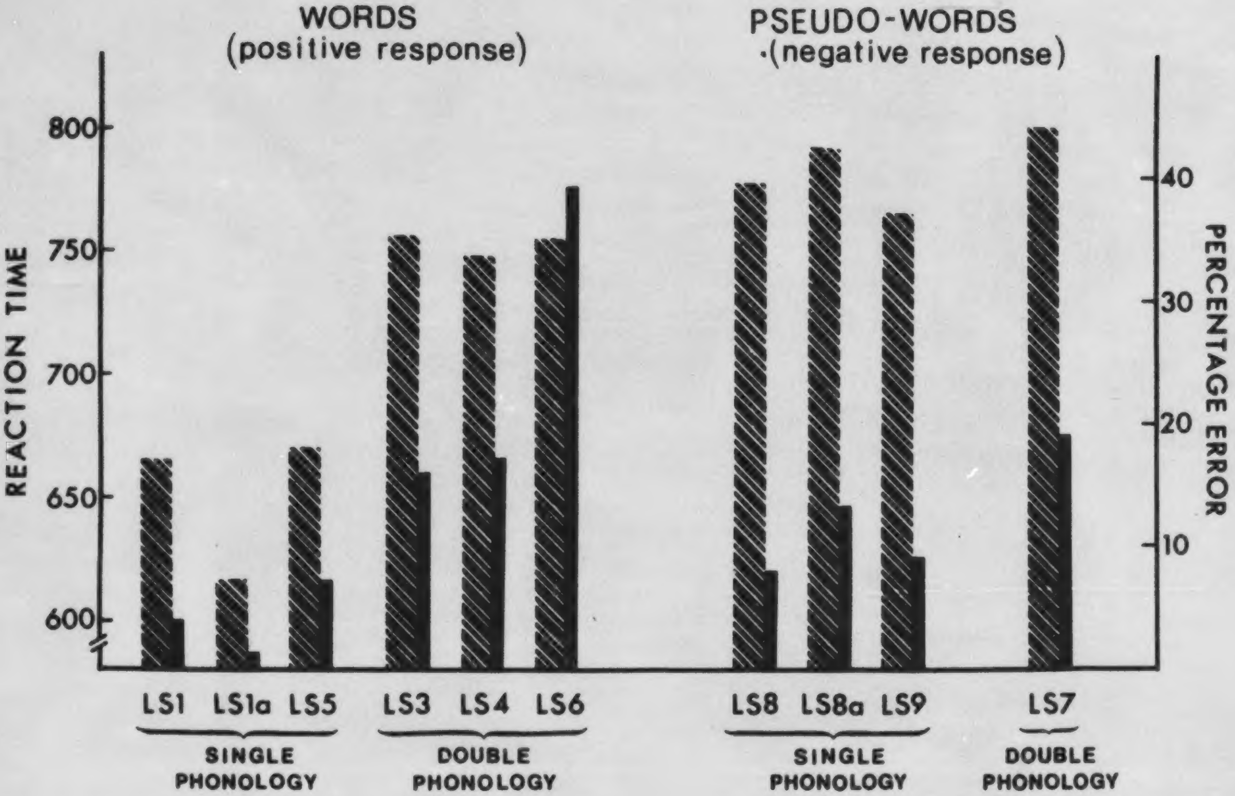


Figure 2. Latencies and errors (too slow and wrong) for lexical decision to 10 types of letter strings. Wide striped bars represent latencies, and thin solid bars represent errors.

of correct reponses for each comparison of interest. Significant differences were found between errors to LS1 and to those of LS4 ($p<.001$), between errors to LS1a and those to LS6 ($p<.001$), and between errors to LS3 and those to LS5 ($p<.001$). In summary, when a word was readable in both alphabets, lexical decision was slowed and errors were increased.

Let us now consider the decision latencies for negative responses. Decision latency was not significantly slower ($p<.05$) for letter strings of Type LS7 than for letter strings of Types LS8, LS8a, and LS9. However, in view of the greater number of slow responses incurred on letter strings of Type LS7 (by a Wilcoxon signed-ranks test, the difference in slow responses between LS7 and LS8 was significant at the .001 level), the data were reanalyzed ignoring the cutoff criterion for slow responses. That is to say, a second analysis was conducted in which a slow response was not replaced by the subject's mean latency but was included in the analysis as a raw datum. On this analysis, decision time for LS7 was significantly slower than decision times for LS8 ($p<.05$) and LS9 ($p<.05$), but not slower than decision time to LS8a ($p<.05$). In short, there is reason to believe that a letter string's affiliation to both alphabets retards negative decision time, a result that is contrary to the observation made in our previous research on bialphabetical lexical decision.

DISCUSSION

Can we take the present experiment as showing that the phonologic form of Serbo-Croatian letter strings contributes significantly to lexical decision? The general sense of the argument for a nonphonologic route to the lexicon is that the reader uses some aspect of the visual appearance of a letter string to directly access its lexical representation.

One fairly representative account of lexical decision is given by Meyer and Ruddy (Note 1). They interpret the relation between the phonological and visual routes to the lexicon as one of competition. A phonologically constrained search of the lexicon is conducted simultaneously with a visually constrained search, and sometimes it is the former search and sometimes it is the latter search that first accesses the target lexical item. When the access is through the phonology and the language is English (or, presumably, an orthographic cognate), a spelling recheck is conducted to insure against judging homophones as words.

For sake of argument, let us suppose that in the present experiment either the direct visual route was more rapid than the phonological route--so that lexical entries were detected more often than not by reference to the word's visual appearance--or the phonologic route was suppressed on grounds of inefficiency. If either supposition were correct, then our subjects should have accepted words readable in both alphabets as rapidly as they accepted words readable in just one alphabet. Given a Serbo-Croatian word such as CAH, which is read differently in the two alphabets but is a word (dream) only in Cyrillic, a lexical search conducted in reference to its visual appearance should have been no slower than the lexical search conducted in reference to the visual appearance of БОЛ, an unequivocal letter string meaning pain. We are reminded, however, that words such as CAH were responded to more slowly and with considerably more error.

Clearly, an appeal solely to the mechanism of direct visual access will be insufficient to account for the present data. Nevertheless, an appeal to some kind of visually related mechanism might work; that is, the data may still be accommodated by a nonphonological interpretation. Suppose that ambiguous letters are specially tagged in memory, and suppose, further, that the realization of an ambiguous character through graphemic analysis always eventuates in a slowing of visually guided search. On both rational and empirical grounds, however, the latter proposal seems unlikely. Presumably, the reason for slowing lexical search is that the circumstances demand that greater than usual care be taken to avoid erroneous responses; thus, pursuant to each unsuccessful visual match, a check might be made on its validity. But the fact that a character is ambiguous in reference to sound cannot be important to the matching process qua visual matching. Character ambiguity in phonetic interpretation cannot increase the possibility of matching error in the domain of visual feature matching, and the detection of ambiguous characters in a letter string, therefore, cannot be proposed as a sensible reason for slowing visual search. An (unreported) observation from our previous search is of importance in this regard. In Experiment 1 of the Lukatela, Savić, Gligorijević, Ognjenović, and Turvey (1978) experiments, the letter strings of Type LS1 sometimes included an ambiguous character. If the presence of ambiguous characters slows lexical search, then the letter strings that included ambiguous characters should have been accepted with the long latencies characteristic of LS3, LS4, LS6, which they were not, and not with the short latencies of LS1, which they were.

Experimental data also permit us to reject a similar argument that takes the common letters as its focus. In the present experiment, for example, letter strings composed of common letters (LS5) were associated with a response pattern (latency and error) that marks them as more closely related to letter strings of Types LS1 and LS1a than to letter strings of Types LS3, LS4, and LS6. There is, however, a more profound reason for rejecting the idea that the presence of common letters slows lexical decision--the simple fact that most vowels are common to the two alphabets, and, therefore, any letter string consistent with the language must contain common letters.

It remains to be seen whether or not other visual coding arguments can be made that differ substantially from the ones given here. For the present, we take the inadequacy of the above graphically based interpretations of the present data to be an indictment against any purely visual account and, indirectly, as support for the inclusion of a phonologically based interpretation. In summary, we claim that the present data are evidence for a phonological mediary in lexical decision. Let us proceed to examine the consequence of this claim and the kind of mechanism needed to explain how phonological bivalence retards lexical decision.

Insofar as the task before the subject was one that, in theory, could have been performed most efficiently by ignoring the phonetic form of the letter strings, it can be argued that phonologic coding is not optional in lexical decision for readers of Serbo-Croatian, or, more conservatively, that it is not a form of coding that the native reader of Serbo-Croatian can easily avoid. Perhaps it is here that a distinction of potential significance can be drawn between the reading of a phonologically deep orthography such as that of English and a phonologically shallow orthography such as that of Serbo-

Croatian: Acquiring a phonologically deep orthography encourages the development of coding options and a sensitivity to linguistic contexts in which individual coding strategies are optimal; by comparison, acquiring a phonologically shallow orthography encourages neither the development of coding options or (axiomatically) a sensitivity to the situations for which they are most appropriate.

It is not our intention in this last remark to claim that access to the lexicon is, for the reader of Serbo-Croatian, exclusively phonological. Rather we intend to express the notion that the cost of automatizing ways of accessing the Serbo-Croatian lexicon other than through the use of the general, transparent, and productive relation between letter patterns and phonetic form probably outweighs the benefits. A mechanism for directly accessing lexical items from some aspects of the visual appearance of letter strings implies a formidable amount of learning about specific stimuli (see Baron, 1977; Brooks, 1977). The long-term benefit of such learning, if successful, is that lexical access might be expedited (Coltheart et al., 1977). Nevertheless, we are presuming that such extensive learning has to be well motivated, and our feeling is that, in this regard, there is little to spur the Yugoslavian reader, given the spelling-to-sound regularity of the Serbo-Croatian orthographies and the efficient and economical reading mechanisms that it makes possible. In terms of a contrast that others (Baron & Strawson, 1976) have found useful, we would expect that fluent readers of Serbo-Croatian would be disproportionately Phoenician (roughly, treat letter strings as alphabetic) in comparison with fluent readers of English who might divide more evenly on the Phoenician-Chinese (roughly, treat letter strings as logographic) dichotomy.

In seeking an account of the effect of bialphabetic letter structure on lexical decision, we pursue a model of lexical decision recently formulated by Coltheart and his colleagues (Coltheart et al., 1977; Davelaar et al., 1978). Their model is essentially an extension of Morton's (1969, 1970) logogen model, and it can be considered as representative of a different class of models from that represented by the Meyer and Ruddy (Note 1) interpretation and described above.

Each word has its own logogen, understood as a memory device that accepts various kinds of information specifying the nature of a letter string. The requisite information is to be found in the letter string itself, in its visual appearance and its phonological structure, and in the context in which the letter string occurs. Each logogen has a certain threshold that is inversely related, over the long term, to the frequency of usage of the word and, over the short term, to the recency of its usage. On this conception, lexical access is equated with the accumulation by a logogen of information to the threshold level. And "search" is equated with the simultaneous accumulation in a number of different logogens of the information that they can accept. In the logogen view, lexical search is parallel in contrast to the serial search that characterizes the model of Meyer and Ruddy (Note 1) (and that of Forster, 1976).

It is reasonably apparent how the logogen view accommodates positive lexical decision, but it is not obvious how it might accommodate the decision that a letter string does not have a lexical entry. For what would reliably justify a "no" response? Surely, it cannot be the fact that at the moment of the decision no logogen has yet reached threshold because, with further delay, a logogen may well do so. To remedy this inadequacy of the logogen account, Coltheart et al. (1977) have proposed that in a lexical decision task the subject makes use of a temporal criterion, a deadline, which is tied to the onset of the individual letter string and is extended as a direct function of the overall level of activation of the logogens following onset. When the (variable) deadline has expired, the subject responds "no."

The two important parameters of the modified logogen model are the logogen threshold and the decision deadline. When lexical decision is slowed by a letter string's affiliation with both Serbo-Croatian alphabets, which of these two parameters bears the responsibility? The arguments of Coltheart et al. (1977) highlight the greater flexibility of the deadline parameter, so let us consider that first. The fact that a letter string of Types LS3, LS4, LS6, and LS7 is phonologically bivalent might mean that the number of logogens such a letter string excites exceeds the number excited by a letter string readable in only one alphabet. This means, on the modified logogen view, that the deadline must be later for phonologically bivalent letter strings. Consider the comparison between LS7, on the one hand, and LS8 and LS8a, on the other. If phonological bivalence extends the deadline, then rejection latencies should be slower for LS7. We recall that the number of responses exceeding our cutoff of 1,300 msec, responses designated as errors, were significantly greater for LS7 than for LS8 and LS8a and, further, that when the latency data were reanalyzed without the cutoff criterion, responses to LS7 were significantly slower than responses to LS8 but not those to LS8a. These results are compatible with an extended deadline interpretation of phonological bivalence. We should note, however, that our previous research (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978) failed to demonstrate an effect of phonological bivalence on negative responses. As remarked at the outset, the present experiment is distinguished from the preceding ones in that no alphabet bias was imposed upon the subjects, and that, in and of itself, may be sufficient reason for the different pattern of results for negative responses. Importantly, however, it is only in this one result that the present and previous experiments differ; in all other outcomes they are virtually identical.

But if it can be agreed that phonological bivalence extends the deadline, how would that fact account for the pattern of results for positive decision? It would be nonsense to assume that positive decisions are delayed until the deadline is reached. While such an assumption correctly predicts slower latencies for words read differently in the two alphabet vs. words readable in only one alphabet, it incorrectly predicts that positive and negative response latencies should be the same. Perhaps we need to consider the possibility that phonological bivalence also influences the threshold parameter. If phonological bivalence raises logogen thresholds across the board, then we would expect positive decisions to be slowed. With the threshold raised more time would be needed to accumulate the evidence sufficient to trigger a logogen.

To effect a raising of threshold that is contingent on a letter string's readability in both alphabets requires a mechanism that monitors the consequences of the graphemic-to-phonemic mapping and adds a constant to the threshold value of each individual logogen on the occasion that two distinct phonologic interpretations arise for a given letter string. The nature of this mechanism is admittedly ad hoc, but then so is the mechanism proposed by Coltheart et al. (1977) to modulate the decision deadline according to the excitation level of the lexicon. But the ad hoc feature of the threshold-raising mechanism is a lesser source of discomfort than is the absence of a rationalization for it.

It would be prudent to raise the thresholds of lexical entries in conditions of stimulation and context that are likely to exaggerate the false alarm rate. Can we argue that the condition of phonological bivalence is such a condition? When interpreting the negative response data, we assumed that when a letter string could receive two distinct phonological descriptions more logogens would be excited than when the letter string was phonologically singular; we assumed, in short, that phonological bivalence delays the deadline. In general, a direct relation between the level of excitation of the internal lexicon and the deadline for negative responses is rational: The more logogens excited, the more likely it is that the proper response is "yes"; if the lexicon is relatively quiescent, the proper response is more likely to be "no." Here, then, is our dilemma. We have said that when a letter string can receive two different phonological interpretations the deadline is extended to guard against misses. The very reasonableness of this statement is argument against the claim that when a letter string can receive two different phonological interpretations, the thresholds are raised to guard against false alarms. We cannot have our cake and eat it too. The benefits of delaying the deadline would be erased by raising the thresholds.

Perhaps we should credit phonological bivalence not with the raising of thresholds but with a slowing down in the process that determines the phonological structure of a letter string. If that process were slowed when a bialphabetic letter string is presented, then the accumulation of phonologic evidence would be retarded and thresholds would be reached at later intervals. This interpretation of the influence of phonological bivalence on positive responses requires no new mechanisms and no ad hoc adjudicating on the benefits and costs of this or that strategy. The question, however, is whether this interpretation does indeed accommodate the data, particularly the pattern of errors. A rough analysis suggests that it does.

Slow responses and incorrect responses were considerably more frequent for words readable in both alphabets than for words readable in just one alphabet. One way to account for the incorrect responses is to suppose that on some occasions the decision deadline was exceeded before a threshold was reached. The slower the determination of the phonological structure of a letter string, the lower the rate at which the level of lexical excitation rises and the longer the period before the deadline undergoes appreciable extension. Consequently, a substantial change in the decision deadline will, on some occasions, not occur rapidly enough to offset the slowed accumulation of phonological evidence, and a "no" response will be emitted.

There is another mechanism that might be proposed that would similarly produce the desired consequence of slowing the rate at which evidence in individual logogens accumulates when the target letter string is readable in two ways. The locus of this alternative mechanism is within the logogen system itself rather than prefatory to it. Specifically, the mechanism is a parallel search procedure of limited power. The operating characteristic of such a search mechanism is that the more representations excited in parallel, the slower the rate at which any individual representation approaches its threshold (Anderson, 1976).

The foregoing considerations of the mechanisms underlying lexical decision are not by any means exhaustive, nor are they intended to be so. At best, they sketch out possible approaches to the data of the present experiment and of those reported previously (Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978). We should not, however, let the difficulty of ascribing a mechanism obscure the conclusion to which the present data point: For the phonologically shallow writing systems of Serbo-Croatian, lexical decision proceeds with reference to the phonology.

REFERENCE NOTE

1. Meyer, D. E., & Ruddy, M. *Lexical memory retrieval based on graphemic and phonemic representations of printed words*. Paper presented at the annual meeting of the Psychonomic Society, St. Louis, November 1973.

REFERENCES

Anderson, J. R. *Language, memory, and thought*. Hillsdale, N.J.: Erlbaum, 1976.

Baron, J. Mechanisms for pronouncing printed words: Use and acquisition. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, N.J.: Erlbaum, 1977.

Baron, J., & Strawson, C. Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 386-393.

Brooks, L. R. Visual pattern in fluent word identification. In A. Reber & D. Scarborough (Eds.), *Towards a psychology of reading*. Hillsdale, N.J.: Erlbaum, 1977.

Chomsky, N. Phonology and reading. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books, 1970.

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI*. London: Academic Press, 1977.

Davelaar, E., Coltheart, M., Besner, D., & Jonasson, J. T. Phonological recoding and lexical access. *Memory & Cognition*, 1978, *6*, 391-402.

Forster, K. I. Accessing the mental lexicon. In R. J. Wales & E. C. T. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam: North-Holland, 1976.

Goodman, K. The 13th easy way to make learning to read difficult: A reaction to Gleitman and Rozin. *Reading Research Quarterly*, 1973, *8*, 484-493.

Kolers, P. Three stages of reading. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books, 1970.

Liberman, I., Liberman, A., Mattingly, I. G., & Shankweiler, D. Orthography and the beginning reader. In J. Kavanagh & R. Venezky (Eds.), *Orthography, reading, and dyslexia*. Baltimore, Md.: University Park Press, 1980.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. Bi-alphabetical lexical decision. *Language and Speech*, 1978, *21*, 142-165.

Lukatela, G., Savić, M., Ognjenović, P., & Turvey, M. T. On the relation between processing the Roman and the Cyrillic alphabets: A preliminary analysis with bi-alphabetical readers. *Language and Speech*, 1978, *21*, 113-141.

Lukatela, G., & Turvey, M. T. Some experiments on the Roman and Cyrillic alphabets of Serbo-Croatian. In J. Kavanagh & R. Venezky (Eds.), *Orthography, reading, and dyslexia*. Baltimore, Md.: University Park Press, 1980.

Lukić, V. [Active written vocabulary of pupils at the elementary school age.] Belgrade: Zavod za Izdavanje Udžbenika SR Srbije, 1970.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Functions of graphemic and phonemic codes in visual word recognition. *Memory & Cognition*, 1974, *2*, 309-321.

Morton, J. Interaction of information in word recognition. *Psychological Review*, 1969, *76*, 165-178.

Morton, J. A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Posner, M. I., & Snyder, C. R. Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition*. Hillsdale, N.J.: Erlbaum, 1975.

Rozin, P., & Gleitman, L. R. The structure and acquisition of reading II. The reading process and the acquisition of the alphabetic principle. In A. Reber & D. Scarborough (Eds.), *Towards a psychology of reading*. Hillsdale, N.J.: Erlbaum, 1977.

Rubenstein, H. R., Lewis, S. S., & Rubenstein, M. A. Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 645-657.

Smith, F. *Understanding reading*. New York: Holt, Rinehart, & Winston, 1971.

FOOTNOTE

[1] It can be argued that for English the representational medium of relevance to the internal lexicon and its access is probably phonological. Thus, any word in the English lexicon is conveyed as a sequence of systematic phonemes divided into its constituent morphemes. For example, "heal" and "health" have the morphophonemic representations /hēl/ and /hēl + θ/. These representations are distinct from their phonetic counterparts; "heal" and "health" are realized approximately as [hīyl] and [hēlθ]. In the phonetic representation of an English word the underlying morphophonemic form is often disguised and the morphophonemic boundaries absent (see Liberman et al., 1980). In contrast with English, we claim here that the phonetic representation of Serbo-Croatian words is virtually indistinguishable from the phonological representation.

REPRESENTATION OF INFLECTED NOUNS IN THE INTERNAL LEXICON*

G. Lukatela+, B. Gligorijević+, A. Kostić++, and M. T. Turvey++

Abstract. The lexical representation of Serbo-Croatian nouns was investigated in a lexical decision task. Because Serbo-Croatian nouns are declined, a noun may appear in one of several grammatical cases distinguished by the inflectional morpheme affixed to the base form. The grammatical cases occur with different frequencies although some are visually and phonetically identical. When the frequencies of identical forms are compounded, the ordering of frequencies is not the same for masculine and feminine genders. These two genders are distinguished further by the fact that the base form for masculine nouns is an actual grammatical case, the nominative singular, whereas the base form for feminine nouns is an abstraction in that it cannot stand alone as an independent word. Exploiting these characteristics of the Serbo-Croatian language, we contrasted three views of how a noun is represented: (1) The independent entries hypothesis, which assumes an independent representation for each grammatical case reflecting its frequency of occurrence; (2) the derivational hypothesis, which assumes that only the base morpheme is stored with the individual cases derived from separately stored inflectional morphemes and rules for combination; and (3) the satellite entries hypothesis, which assumes that all cases are individually represented with the nominative singular functioning as the nucleus and the embodiment of the noun's frequency and around which the other cases cluster uniformly. The evidence strongly favors the satellite entries hypothesis.

Inflection is the major grammatical device of Serbo-Croatian, Yugoslavia's principal language. In general, the grammatical cases of nouns are formed by adding a suffix to a root morpheme where the suffix is of the vowel or vowel-consonant or vowel-consonant-vowel type. Less frequently, inflection involves additional processes such as vowel deletion and consonant palatalization.

The grammatical cases of Serbo-Croatian nouns produced by inflection are not equal in their frequency of occurrence. Table 1 summarizes the frequency analysis of Dj. Kostić (1965) on a corpus of approximately two million Serbo-Croatian words appearing in the daily press and contemporary poetry. The non-

---

*To appear in Memory & Cognition.

+University of Belgrade.

++Also University of Connecticut.

Acknowledgment. This research was supported in part by NICHD Grant HD-08495 to the University of Belgrade and in part by NICHD Grant HD-01994 to Haskins Laboratories.

Table 1

Case frequencies in percentages*

	Singular				Plural			
	Masculine	Feminine	Neuter	Total	Masculine	Feminine	Neuter	Total
Nominative	12.83	8.84	2.88	24.55	3.33	3.58	0.69	7.60
	(28.89)	(22.56)	(20.44)		(7.50)	(9.14)	(4.30)	
Genitive	8.56	7.88	3.47	19.91	3.96	3.22	0.61	7.79
	(19.27)	(20.11)	(24.63)		(8.92)	(8.22)	(4.33)	
Dative	0.87	0.38	0.31	1.56	0.28	0.16	0.04	0.47
	(1.96)	(0.97)	(2.20)		(0.63)	(0.41)	(0.28)	
Accusative	5.49	5.48	2.55	13.52	2.21	2.75	0.73	5.69
	(12.36)	(13.99)	(18.10)		(4.98)	(7.02)	(5.18)	
Instrumental	1.90	1.94	0.86	4.70	0.60	0.80	0.13	1.46
	(4.28)	(4.95)	(6.10)		(1.35)	(1.86)	(0.92)	
Locative	3.77	3.42	1.61	8.80	0.61	0.80	0.21	1.62
	(8.48)	(8.73)	(11.43)		(1.37)	(2.04)	(1.48)	
TOTAL	33.42	27.94	11.68	73.04	10.99	11.24	2.41	23.64
	(75.25)	(71.31)	(82.89)		(24.75)	(28.69)	(17.11)	

*Table is adopted from Dj. Kostić (1965). Figures in parenthesis represent the normalized percentages as related to the particular gender. Percentages do not add to 100 percent owing to the omission of the rarely occurring vocative case.

parenthesized numbers are actual percentages. Thus for all nouns in the corpus, 12.83 percent were masculine nouns in the nominative singular, 7.88 percent were feminine nouns in the genitive singular, 0.13 percent were neuter nouns in the instrumental plural, and so on. Reading the totals, we see that most nouns were masculine and that the nominative singular was the most popular grammatical case. The parenthesized numbers are normalized percentages and can be read as follows, taking the masculine gender as an example. For any given masculine noun that occurs in the language with frequency _f_, the nominative singular form of that noun occurs with a frequency of approximately .29_f_, the genitive singular form with a frequency of approximately .19_f_, the dative singular form with a frequency of approximately .02_f_, and so on. In short, the normalized percentage for a given grammatical case of a given gender is the likelihood that when a noun of that gender appears, it appears in that particular case.

The question of interest to the present paper is how the inflected Serbo-Croatian nouns are represented in lexical memory. Following MacKay (1978) and Manelis and Tharp (1977), we can distinguish two hypotheses about the lexical representation of words with common morphological stems. According to the independent entries hypothesis, the individual grammatical forms of a Serbo-Croatian noun would be represented in the lexicon by independent representations, one internal representation for each grammatical form. On the derivational hypothesis, rather than instantiating all the forms of a given noun in the internal lexicon there would be but one instantiation, probably of the noun's root morpheme. There would also be in memory only a single instantiation of the set of inflectional morphemes. Appropriate combinations of the root morpheme and inflections would be determined by separately stored syntactic rules.

There have been relatively few direct contrasts of the two hypotheses for English lexical items and the results have been largely equivocal. Manelis and Tharp (1977) compared lexical decision ("Is this letter string a word?") times for pairs of affixed words (words consisting of two morphemes, a root morpheme and a suffix) with lexical decision times for pairs of nonaffixed words (words consisting of a single morpheme). Manelis and Tharp (1977) predicted two possible outcomes from the derivational or, as they termed it, decompositional hypothesis. For a given letter string, decomposition into root and ending could be an obligatory first step with lexical search for the whole item a contingent later step; or, lexical search for the whole item could be the initial obligatory step with decomposition occurring later and dependent upon failure to find the whole item in memory. Consider the prediction that follows from the notion that decomposition occurs first. A word--whether it be affixed or nonaffixed--is partitioned into root and ending. A test is then made to determine the validity of the combination as an affixed word. If the combination proves valid, a positive response is initiated; if it proves invalid (meaning that the word is nonaffixed), a search of the lexicon is conducted for the nondecomposed letter string. In brief, with everything else equal, the decomposition-first argument predicts faster lexical decision for affixed words than for nonaffixed words. The contrary prediction follows from the decomposition-second argument. If the initial search of the lexicon for the nondecomposed letter string is successful (meaning that the letter string is a nonaffixed word), then a positive response can be initiated. However, if the search is unsuccessful, then the letter string is decomposed and the combination of root and affix tested for its validity. Obviously, on the decomposition-second argument, lexical decision should be slower for affixed words. The Manelis and Tharp (1977) investigation failed to find a difference between affixed and nonaffixed words in either direction, a result that favored the independent entries hypothesis over either version of the decompositional hypothesis.

However, the failure to find evidence for morphological decomposition with suffixed words contrasts with the provision of such evidence by Taft and Forster (1975) for prefixed words. These investigators reported that rejecting real roots (for example, SULTS as in INSULTS) as words took longer than rejecting false roots (for example, NINGS as in INNINGS) as words. The interpretation given was that real stems would be found in the lexicon and a subsequent check would be needed to determine that these lexical entries do not constitute words in the absence of an appropriate prefix.

A further demonstration of morphological decomposition is reported by MacKay (1978), although his experiment is distinguished from the experiment described above in that it looks at the production of words rather than at their perception. Subjects heard verbs (for example, conclude, decide) that they had to nominalize (conclusion, decision) as rapidly as possible (MacKay, 1978). It was shown that certain nominalizations took longer than others, precisely, the more complicated the derivational process--the more steps intervening between verb form and noun form--the slower the nominalizations.

The source of the discrepancy between the experiments of Manelis and Tharp (1977) and MacKay (1978) could be relatively trivial--a matter of differences in methodology. On the other hand, the discrepancy might arise from a deep-seated difference between the kind of memory structure needed to recognize words and the kind of memory structure needed to produce them. In the former case the analogy that has come to be adopted is that of a dictionary: The internal representations of words are coded on orthographic and phonological principles and are accessed accordingly. But in the latter case--that of the requirements of production--the opposite analogy is not that of a dictionary but of a thesaurus (Labov, 1978): The internal representations of words are coded on semantic principles and should be accessed accordingly; for in production the problem is to locate a word that expresses a given meaning.

Whatever the reason for the equivocality identified above we should note that, with regard to the representation of inflected nouns, the independent entries hypothesis and derivational hypothesis are not exclusive. A third hypothesis can be entertained, which combines features of the first two. We refer to it, picturesquely, as the "satellite" entries hypothesis. Here are its distinguishing characteristics: (1) each grammatical case of a noun has a separate entry in the lexicon; (2) the nominative singular entry functions as the nucleus of the noun and it expresses the frequency of occurrence of the noun that it represents; (3) lexical entries of the remaining grammatical cases cluster (relatively) uniformly about the nominative singular entry and are organized among themselves and in relation to the nominative singular by a (for now unspecified) principle other than frequency. In short, the lexical entries of the oblique cases of a noun are satellites to the lexical entry of the noun's nominative singular.

The second characteristic of the satellite entries hypothesis reflects a common assumption of hypotheses about lexical memory, namely, that entries in the lexicon express the frequency of the word they represent. We pursue that assumption in the remarks that follow because it figures significantly in the eventual predictions we wish to make.

There are two fashionable interpretations of how a word's frequency of occurrence is coded in the internal lexicon. The entries in lexical memory may be likened to the files in a filing cabinet ordered according to frequency of usage (Forster & Bednall, 1976; Rubenstein, Lewis, & Rubenstein, 1971; Stanners & Forbach, 1973). A word's frequency of occurrence is expressed in lexical memory by the location of its lexical entry. Thus, on the filing-cabinet analogy, the entries for the most frequently occurring words are to be found at the front of the cabinet (at the start of lexical search) while those entries for the least frequently occurring words are to be found at the back of the cabinet (toward the end of lexical search). On this view, lexical search is serial and its duration is inversely related to the frequency of occurrence of the target word; when no lexical entry is to be found--that is, when the letter string is a nonword--the search is exhaustive. If the filing-cabinet account of the coding of word frequency in lexical memory can be referred to as an inter-entry account, then its popular alternative can be referred to as an intra-entry account, for here the emphasis is not on an entry's position relative to other entries but on the individual entry's sensitivity to linguistic stimulation. According to the intra-entry account each lexical entry is a device for accepting evidence about the presence of the word it represents (see the logogen model of Morton, 1969, 1970). In the case where the word in question occurs very frequently, the evidence needed for detecting its presence is less or, equivalently, the threshold of its lexical entry is lower, than in the case where the word in question occurs rarely. On this view, lexical search is parallel and, in common with the inter-entry view, its duration is inversely related to a word's frequency of occurrence. It is not so clear, however, how the intra-entry view accounts for decision time when no lexical entry is to be found (see Coltheart, Davelaar, Jonasson, & Besner, 1977).

If there is an independent entry for each grammatical case of a Serbo-Croatian noun, then we might suppose that lexical decision times for the grammatical cases of a given noun will vary in proportion to their frequencies of occurrence. In a previous experiment (Lukatela, Mandić, Gligorijević, Kostić, Savić, & Turvey, 1978), we examined this prediction from the independent entries hypothesis and found it wanting. Lexical decision time was not related by a unique, constant multiplier to the corresponding logarithms of the proportional frequencies of three grammatical cases. Rather, the decision time for one case, the nominative singular, was significantly less than the decision time to either of the other two cases (the instrumental singular and the dative singular), which did not differ one from the other in terms of decision time even though they differed in frequency. We interpreted this observation as support for either a derivational hypothesis or a hypothesis consonant with the point of view that the nominative singular is the nucleus entry about which the entries for the other grammatical cases cluster uniformly.

The experiment to be reported here contrasts the satellite-entries hypothesis with the independent entries hypothesis on the one hand and with the derivational hypothesis on the other. To anticipate, the outcome of the experiment favors the satellite entries interpretation of the lexical organization of inflected nouns.

The experiment takes advantage of two facts of the Serbo-Croatian language. First, the same letter pattern (and, therefore, phonetic pattern) can represent more than one grammatical case. For example, the inanimate noun ŠERPA (nominative singular form), which means *pot*, is written as ŠERPE and pronounced identically in the genitive singular, nominative plural and accusative plural. Where identities exist, the case frequencies can be compounded. The case identities and their compound frequencies for nouns of the masculine and feminine genders are given in Table 2.

The second fact to be exploited is that whereas the nominative singular is the root morpheme in the declension of masculine nouns, it is not the root morpheme in the declension of feminine nouns. For the latter the root morpheme is an abstraction in the loose sense that the root morpheme never occurs as an actual grammatical case. In terms of distinctions sometimes used by linguists, the root morpheme of masculine nouns is *full* (it has semantic content) and *free* (it can stand alone as an independent word), whereas the root morpheme of feminine nouns is less obviously full and it is certainly not free. Table 3 gives examples of the two genders.

Let us return to the first fact identified above and put it to use as a means of prying apart the perspective of satellite entries from that of independent entries. The compounded frequency of the nominative singular form in the masculine gender proves to be *greater* than that of the genitive singular form in the masculine gender. For nouns of the feminine gender this relation is reversed: the nominative singular form occurs *less* frequently than the genitive singular. Thus, for a masculine noun of frequency of occurrence f, the respective proportional frequencies of the nominative singular and genitive singular letter patterns are approximately $.41f$ and $.28f$. In contrast, for a feminine noun of frequency of occurrence f, the respective proportional frequencies are approximately $.31f$ and $.36f$. The independent entries hypothesis would predict a shorter latency lexical decision for nominative singular masculine nouns than for genitive singular masculine nouns. That same hypothesis, however, with respect to feminine nouns would predict either little difference in lexical decision latency for the two grammatical cases or a difference in which the decision time to the genitive singular form is the briefer of the two. In comparison the satellite entries hypothesis makes a considerably simpler prediction: For both genders the nominative singular will be responded to faster than the genitive singular.

The two hypotheses can be further contrasted with respect to their predictions on lexical decision times to the instrumental singular, which occurs with a proportional frequency of approximately $.04f$ in the masculine and approximately $.05f$ in the feminine. The independent entries hypothesis would predict that decision times to the very low frequency instrumental singular of both genders should be much longer than the decision times for the high frequency nominative singular and the high frequency genitive singular.

Table 2

Identical grammatical cases and their compound frequencies

Masculine nouns (inanimate)	Percent Occurrence	Feminine nouns	Percent Occurrence
Nominative singular, accusative singular	41.25	Nominative singular, genitive plural	30.78
genitive singular, genitive plural	28.19	Genitive singular, nominative plural, accusative plural	36.27
Locative singular, dative singular	10.45	Locative singular, dative singular	9.70

Table 3

Declension of a masculine noun and of a feminine noun

Case	Masculine		Feminine	
	Singular	Plural	Singular	Plural
Nominative	DINAR (money)	DINARI	ŽENA (woman)	ŽENE
Genitive	DINARA	DINARA	ŽENE	ŽENA
Dative	DINARU	DINARIMA	ŽENI	ŽENAMA
Accusative	DINAR	DINARE	ŽENU	ŽENE
Vocative	DINARE	DINARI	ŽENO	ŽENE
Instrumental	DINAROM	DINARIMA	ŽENOM	ŽENAMA
Loctive	DINARU	DINARIMA	ŽENI	ŽENAMA

The satellite entries hypothesis, in contrast, predicts that lexical decision time for the instrumental singular should, in both genders, be very close--probably identical--to that for the genitive singular and significantly longer than that for the nominative singular. A summary of these contrasting predictions of the two hypotheses is given in Table 4 where the inequality symbols are in reference to lexical decision time and the letters identify the nominative singular (ns), genitive singular (gs) and instrumental singular (is).

The rationale for pooling the frequencies of visually identical cases is that a reader's sensitivity (in lexical decision) to a given grammatical form of a given noun is determined solely by the relative frequencies with which the reader has seen that grammatical form as a visual object. A different perspective, however, and one that is more consonant with the satellite-entries hypothesis, is that it is the visual form in a sentential context--that is, as a grammatical object rather than as a crass visual object--that is important so that there are indeed separate lexical entries for individual cases that are visually identical but grammatically distinct. On this latter perspective we should predict latency relations on the basis of the uncompounded frequencies as given in Table 1. The relevant predictions are shown in Table 5 and, as comparison of Tables 4 and 5 reveals, the predictions from compounded and uncompounded frequencies differ only slightly.

Let us now take the second fact identified above, namely, the differential status of the nominative singular in nouns of the masculine and feminine gender, and put it to use for the purpose of distinguishing the satellite entries perspective from that of derivation. Recalling the Manelis and Tharp (1977) analysis, in lexical decisions an affixed word would be decomposed into base morpheme and affix and the combination then evaluated for its validity. Consider this derivational account of lexical decisions as applied to the grammatical cases of masculine and feminine nouns exemplified in Table 3. The base morpheme of the masculine noun in Table 3 is DINAR, which is also the nominative singular, but the base morpheme of the feminine noun is ŽEN, which is not identical with any grammatical case. By one reading of the derivational account of lexical decisions, the decision process for the feminine nominative singular ŽENA should differ from that for the masculine nominative singular DINAR. Since ŽEN and not ŽENA is represented in memory, ŽENA would have to be decomposed into the two morphemes ŽEN and A and the combination then assessed for its validity. Therefore, whether decomposition occurs before or after lexical search, the decision process for ŽENA should not differ from the decision processes for the other grammatical cases, which similarly are decomposable into the root ŽEN and a single inflectional morpheme. But consider the relation between DINAR and its allied oblique cases. If lexical search for the whole unit preceded decomposition, then DINAR's lexical legitimacy would be determined in the first state but the determination of (say) DINAROM's lexical status would have to await the second stage. On the decomposition-second version of the derivational view, decision times for the nominative singular of masculine nouns should be shorter than those for the grammatical cases that are inflected and that, in turn, should not differ among themselves. However, if decomposition precedes lexical search, then a different outcome is to be expected. In comparison to the oblique cases, DINAR would resist sensible decomposition and would have to be processed through the subsequent stage of lexical search--in which case

Table 4

Predictions of independent entries and satellite entries hypotheses for compounded frequencies.

Hypothesis	Masculine nouns	Feminine nouns
Independent entries	$ns < gs < is$	$ns \geq gs < is$
Satellite entries	$ns < gs = is$	$ns < gs = is$

Table 5

Predictions of independent entries and satellite entries hypothesis for uncompounded frequencies

Hypothesis	Masculine nouns	Feminine nouns
Independent entries	$ns < gs < is$	$ns \leq gs < is$
Satellite entries	$ns < gs = is$	$ns < gs = is$

lexical decision to the nominative singular would be the slowest, not the fastest.

There is yet another possibility. When DINAR is subjected to the decomposition stage, the decomposition process yields two morphemes, DINAR and the null morpheme, ϕ, which are then assessed as constituting a legal combination. As a modification of the decomposition-first argument, this latter argument predicts no difference in lexical decision times among the grammatical cases of masculine nouns.

Table 6 summarizes the contrasting predictions of the derivational and satellite-entries hypotheses. The important thing to note is that the satellite-entries view differs from the decomposition-first and decomposition-second views in that it predicts the same pattern of latencies for masculine and feminine nouns and from the modified decomposition-first view in that it predicts a difference among grammatical cases. It remains for us to point out that differences between the derivational and satellite-entries hypotheses remain even if the frequency factor is incorporated into the predictions of the three versions of the derivational hypothesis. Borrowing a strategy popular with writers of mathematics textbooks, we leave the generation of these predictions as an exercise for the reader.

Table 6

Predictions of derivational and satellite entries hypotheses

Hypothesis	Masculine nouns	Feminine nouns
Decomposition second	ns < gs = is	ns = gs = is
Decomposition first	ns > gs = is	ns = gs = is
Modified decomposition first	ns = gs = is	ns = gs = is
Satellite entries	ns < gs = is	ns < gs = is

Method

Subjects

Sixty undergraduate students from the Psychology Department of the University of Belgrade participated in the experiment. All subjects had had previous experience with reaction time experiments. Some of the subjects had participated in lexical decision experiments before, but none had done so within a month of the present experiment. Moreover, few of the words of the present experiment had been used in the earlier experiments.

Materials

Twenty-seven feminine nouns and twenty-seven masculine nouns were selected according to the following criteria: (1) all the nouns had to be easily imagined, that is, they had to be concrete nouns; (2) all the nouns had to be easy to read aloud in all grammatical cases, that is, consonant runs were avoided: (3) all the nouns had to have only a single meaning invariant over grammatical cases; (4) all the nouns had to be regular; and (5) all the masculine nouns had to be inanimate. Nouns that met these criteria were equated in frequency of occurrence (Lukić, 1970).

Three 35-mm slides were constructed for each noun: one for the noun's nominative singular, one for the noun's genitive singular and one for the noun's instrumental singular. Accordingly, there was a total of 162 slides in which the string of Roman (see Lukatela, Savić, Ognjenović, & Turvey, 1978) letters (Helvetia light, 12 point), arranged horizontally at the center of the slide, spelled a word in Serbo-Croatian.

A set of 162 pseudoword slides was constructed by converting a different list of words meeting the same criteria as above into a pseudoword. This was done in the nominative singular and genitive singular cases by changing the first letter and in the instrumental singular case by changing the last letter so as to avoid idiosyncratic instrumental endings.

Procedure

On each trial, the subject's task was to decide as rapidly as possible whether the presented letter string was a word or a pseudoword. Each slide was exposed for 1500 msec in one channel of a three-channel tachistoscope (Scientific Prototype, Model GB) illuminated at 10.3 cd/m^2. Both hands were used in responding to the stimuli. Both thumbs were placed on a telegraph key button close to the subject and both forefingers on another telegraph key button two inches further away. The closer button was depressed for a "No" response (the string of letters was not a word), and the further button was depressed for a "Yes" response (the string of letters was a word).

Latency was measured from stimulus onset. The total session lasted for half an hour with a short pause after every eighteen slides.

Design

Each subject saw a total of 108 slides of which 54 were words and 54 were pseudowords, but no subject saw any given letter string or any given noun more than once in the course of the experiment. This was achieved in the following manner. The 54 feminine and masculine nouns were divided into three groups (A,B,C) of 18 nouns each. The sixty subjects were divided into three groups (1,2,3) of 20 subjects each. Subjects in Group 1 saw the nominative singular cases of category A nouns, the genitive singular of category B nouns and the instrumental singular of category C nouns. Subjects in Group 2 saw the nominative singular case of category B nouns, the genitive singular of category C nouns and the instrumental singular of category A nouns. For subjects in Group 3 the categories were C, A, B, respectively, for nominative, genitive and instrumental. A similar partitioning into categories and mapping onto subject groups was done for the pseudowords.

Results

Figure 1 gives a histogram plot of the mean reaction times for the three grammatical cases of the masculine and feminine nouns. Reaction times less than 300 msec and greater than 1500 msec were excluded from the calculations of the means, as were erroneous responses that occurred in the present experiment at a rate of less than 2.5 percent. Only the latencies to words are considered in the analysis below.

Inspection of Figure 1 suggests a difference in the rank order of grammatical-case latencies between genders. At the same time, however, the figure does not suggest a pattern of results consonant with the predictions of the alternatives to the satellite-entries hypothesis. A difference between the genders might hold for the absolute latencies. The apparently slower overall response to the masculine nouns might be owing to their generally greater length in both number of letters and number of syllables. Word length is known to contribute significantly to response latencies (Whaley, 1978).

The design of the present experiment was chosen to insure that no subject saw the same noun twice. It is a design, however, that raises certain difficulties where one is concerned with keeping the analysis true to the strictures advocated by Clark (1973), that is, of treating both subjects and letter strings as "random effects" and computing reliability of results over both of these sampling domains. To circumvent these difficulties we use a variation of a procedure that we have reported previously (see Lukatela, Savić, Gligorijević, Ognjenović, & Turvey, 1978).

A comparison within a gender between any two of the three grammatical cases is composed of two subcomparisons: one in which the nouns are the same, but the subjects are different (comparing decision times for A words, B words and C words) and one in which the subjects are the same, but the nouns are different (comparing decision times for Group 1, Group 2 and Group 3). The two quasi-F ratios for these subcomparisons are viewed as random variables the probabilities of which have a Chi-square distribution with 2 x 2 degrees of freedom. These new random variables are computed as $r_i = -2 \ln (p_i)$ for any subcomparison r_i for which the F' is at the probability level p_i. The obtained sum of the new variables is then assessed for significance against the Chi-square value for the corresponding degrees of freedom. In short, this analysis assesses the likelihood that a set of two quasi-F ratios with probabilities of p_1, p_2 could have come about by chance.

For the masculine nouns the nominative singular differed from both the genitive singular, $\chi^2(4) = 28.65$, $p < .001$, and the instrumental singular, $\chi^2(4) = 19.44$, $p < .001$, which did not differ between themselves, $\chi^2(4) = 5.51$, $p > .05$. The same pattern holds for the feminine nouns: nominative singular vs. genitive singular, $\chi^2(4) = 29.46$, $p < .001$, nominative singular vs. instrumental singular, $\chi^2(4) = 35.45$, $p < .001$; genitive singular vs. instrumental singular, $\chi^2(4) = 1.58$, $p > .05$.

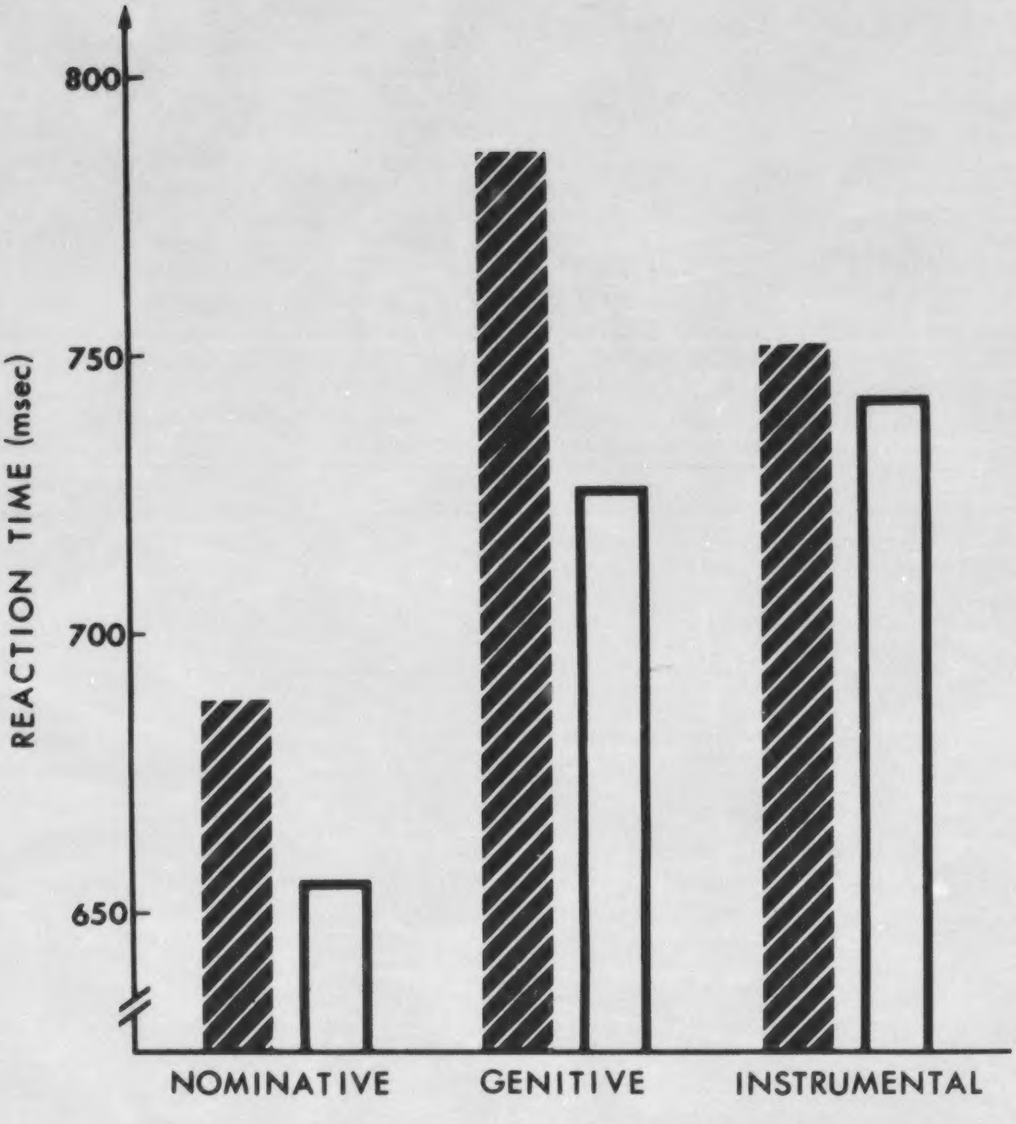


Figure 1. Reaction time to three grammatical cases of nouns of the masculine gender (striped bars) and nouns of the feminine gender.

Discussion

The purpose of the present experiment was to assess three interpretations of how the inflected nouns of the Serbo-Croatian language are represented in the internal lexicon. On one interpretation, the independent-entries hypothesis, it is assumed that each grammatical case is stored in the lexicon as a separate and relatively independent entry. Insofar as an entry in the internal lexicon is believed to embody--either through its relation to the other entries or through its sensitivity to linguistic stimulation--the frequency of occurrence of the word that it represents, then it should be argued that the grammatical cases of any given noun must relate among themselves in terms of their frequencies of occurrence. This prediction of the independent units hypothesis was examined through an investigation of lexical decision to three grammatical cases--the nominative singular, the genitive singular and the instrumental singular. The relation between the first two cases differs as a function of noun gender: For masculine nouns the nominative singular is of greater compounded frequency, whereas for feminine nouns the genitive singular is (on compounding identical grammatical cases) the more frequently occurring form. In both genders the instrumental singular occurs far less frequently than the other two. The pattern of lexical decision latencies to be expected from the independent units hypothesis was not realized; rather than there being one pattern for the masculine nouns and another for the feminine nouns there was a single pattern, the same for both genders. Importantly, lexical decision time was briefest for the nominative singular of both genders and there was no latency difference between the genitive singular and instrumental singular of both genders.

The obtained results are consistent, therefore, not with an independent-units hypothesis as we have interpreted it, but with a hypothesis that assumes that not all grammatical cases are qualitatively alike in lexical status and that the grammatical cases are not ordered among themselves according to frequency of occurrence. One grammatical case, the nominative singular, appears to play a pivotal role owing in part, perhaps, to its primacy in acquisition (Carroll & White, 1973a, 1973b). The latter fact is important in another way too: it argues against a derivational hypothesis in which lexical decision involves successive stages of decomposing into the root and inflectional morphemes and testing the combination for its legality. Morphologically, the nominative singular of feminine nouns is like all other cases in that it consists of a root form and an inflectional ending, but the nominative singular of masculine nouns is unlike other cases in that it *is* the root form and contains no inflectional ending. Two versions of the derivational hypothesis (see Table 6) predict differences between masculine and feminine nouns in the pattern of decision latencies among the grammatical cases. The experiment revealed, however, that the pattern for the two genders is the same, not different. A third version of the derivational hypothesis does predict identical patterns for masculine and feminine nouns but the predicted pattern is one in which there are no latency differences among grammatical cases. We are reminded that for both genders the experimental outcome was a latency difference that favored the nominative singular over the other two cases. Thus the third version of the derivational hypothesis does not hold either.

Before we draw any general conclusions from the present data, it behooves us to consider an aspect of the design that might give reason for caution. The basis for the fifth restriction on the choice of words described above, that the masculine nouns be inanimate, was that in the declension of nouns of the masculine gender the grammatical cases that are visually/phonologically identical are not the same for nouns denoting animate and inanimate objects. For example, the genitive singular is identical in form to the genitive plural in the case of inanimate nouns and identical in form to the genitive plural and accusative singular in the case of animate nouns. For the compounding of frequencies it seemed prudent to stay with just one kind of masculine noun although either kind would have been adequate for the purposes of the experiment. However, in retrospect, our choice to consider only one of the two kinds of masculine noun may have introduced an unnecessary complication. A native speaker of English unfamiliar with Serbo-Croatian might intuit that the contribution of the animate and inanimate nouns to the relative frequencies of masculine grammatical cases given in Table 1 is not the same (for example, one kind of masculine noun might contribute more to the frequency of one case than to another) and, therefore, to select one of the two kinds of masculine nouns is to make void the use of the tabulated frequencies.

In English, possession is marked by 's. If this form is taken as the sole representative of the genitive case, then given that the use of 's tends to favor animate over inanimate nouns, one might suppose that the genitive case is the hallmark of animate nouns. However, English combines inanimate nouns with the preposition of to produce effectively a partitive genitive--"...of the car," "...of the paper" (see Jaspersen, 1962). It is unlikely that these two kinds of genitives differ markedly in their frequencies of occurrence. In Serbo-Croatian the genitive case, unlike its counterpart in English, is a very complex case assuming thirteen different grammatical functions--of these functions one is exclusively related to animate nouns and three are exclusively related to inanimate nouns (Stefanović, 1974). As with English it seems unlikely that the frequency of the genitive case in Serbo-Croatian would be significantly less for inanimate nouns than for animate nouns.

Similar comments need to be made in reference to the instrumental case, for here one might suppose that inanimate nouns take the instrumental form more so than animate nouns. In Serbo-Croatian there are three categories of instrumental: Instrumental case without preposition (eight kinds); instrumental case with the preposition with (three kinds); and instrumental case with spatial prepositions (above, under, in front of, between/among). Of these three types only two kinds are exclusively related to inanimate nouns (Ivić, Note 1).

Of course, the point we are trying to establish is that the case frequencies for masculine nouns as reported in Table 1, and on the basis of which we formed our predictions concerning the respective hypotheses of lexical organization, are equally applicable to masculine nouns of both the inanimate and animate kind. Nevertheless, in the absence of case frequency norms for individual words (which are not currently available) there is still some room for doubting--although we believe it to be small--that the foregoing contention holds. A small empirical point in our favor is that the mean decision times of thirty-nine subjects for ten animate and ten inanimate masculine nouns drawn from the stimuli of the previous experiment (Lukatela et al., 1978) were virtually identical for both nominative singular and instrumental singular cases: 594 msec and 680 msec, respectively, for the ten inanimate nouns and 591 msec and 674 msec, respectively, for the ten animate nouns. If animate and inanimate masculine nouns differ markedly in the frequency with which they occur in the instrumental case and if decision latency reflected that frequency distinction, then the lexical decision times should have differed.

We would argue, therefore, that taken collectively the present experiment and the previous one (Lukatela et al., 1978) support the assumption that the oblique non-nominative singular cases do not differ in relative accessibility owing to their differences in frequency of occurrence but rather that they are equally accessible. To date we have found little evidence for a difference in lexical decision latencies among the genitive singular, locative singular and instrumental singular cases (and, therefore, in addition, among their visually identical mates, see Table 2).

Suppose that after Morton's (1969, 1970) logogen model we assume that the lexical representation of the nominative singular has a threshold inversely proportional to the frequency with which the noun (indifferent to its particular grammatical case) occurs in the language. Then, given the preceding observation, we should suppose that there is a common threshold level for the logogens of the oblique cases that is at a value equal to the threshold of the nominative singular's logogen incremented by a constant. It is, perhaps, in some such sense as this--in the way in which the thresholds of the lexical entries for oblique grammatical cases are tied by a constant to the threshold of the lexical entry for the nominative singular--that we can begin to interpret the intuitive notion of a satellite organization for the inflected nouns of Serbo-Croatian. In view of the outcome of the present experiment we would conclude that the hypothesis of a nucleus logogen representing the nominative singular and about which the logogens of the oblique cases cluster uniformly is a better candidate for understanding the lexical organization of inflected nouns than either the hypothesis that the cases are represented independently of one another or the hypothesis that they are derived by rule.

Recently (and subsequent to the design and implementation of the present experiment) a description of lexical organization has been proposed (Taft, 1979a) that accommodates the features of both the independent entries and the decomposition hypotheses. The lexicon is said to consist of a master file and a number of peripheral files: orthographic, phonological and semantic (Forster, 1976). In the master file the surface form of each word is separately and completely represented. In the peripheral files, on the other hand (of which the orthographic is the one of special significance to visual word recognition), it is base forms that are represented rather than surface forms. Peripheral files store information that is sufficient for selectively and successfully accessing the master file where all information is to be found. It is argued that in the orthographic file the first syllable of a word, defined orthographically and morphologically, identifies the base form (Taft, 1979b); and that the frequency of a given base form is defined by the summed frequencies of the individual words of which it is the first syllable (Taft, 1979a). Importantly, in both kinds of file, master and peripheral, the frequency of an entry is a significant determinant of access time.

Consider the lexical representation of an inflected Serbo-Croatian noun from the perspective of the master file/peripheral file notion. There would be for a given noun a single entry in the orthographic file—say, the first syllable—with a frequency determined by the noun's occurrence in the language and fourteen entries in the master file (one entry for each grammatical case) with their individual frequencies determined by the frequency of occurrence of the individual cases that they represent. Given nouns such as ŽENA and DINAR, the peripheral file would contain ŽEN and DIN, respectively, whereas the master file would contain, for each of the two nouns, the full form of each grammatical case. Lexical decision occurs via these steps. First, the noun is decomposed into the first syllable and affixes. Second, a search of the peripheral file is conducted for a length of time determined by the frequency of the base form. And third, the master file is accessed (through the address given by the base form entry in the peripheral file) and the legality of the base form/affix(es) combination ascertained at a speed determined by the frequency of the combination (that is, by the frequency of the individual grammatical case). We see, in short, that although the master file/peripheral file notion ascribes to the decomposition hypothesis, it predicts the same outcome as the independent entries hypothesis, namely, that decision times are a function of the relative frequencies of the individual grammatical cases.

Our conclusion concerning the organization of inflected Serbo-Croatian nouns, based as it is on the indifference of decision latency to grammatical case frequency, does not concur with the master file/peripheral file notion—at least not with the current form of the notion, for there are hints that distinct files are a needed conception for certain aspects of lexical access (e.g., Forster, 1979; Glanzer & Ehrenreich, 1979) and, therefore, we would expect the general idea to receive further attention and to undergo modification. One major reason for the lack of concurrence may rest with the issue of whether lexical organization is uniform or pluralistic. Chomsky (1970) and others (e.g., Stanners, Neiser, Hernon, & Hall, 1979) have expressed a pluralistic view, arguing, for example, that the lexicon's organizational formats for the inflectional forms of English verbs and for the nominal derivations of English verbs need not be identical. And Bradley (1978) has given good empirical reasons for holding distinct the lexical organizations of the closed set of words (often termed function words) from the open set of words. Thus, the fact that the affixed English nouns and verbs studied by Taft (1979a) and the inflected Serbo-Croatian nouns studied by us submit to different explanatory accounts of lexical organization may point less to an opposition of data than to a differentiation of lexical organization according to differences in linguistic forms and functions.

REFERENCE NOTE

1. Ivić, M. *Case system in Serbo-Croatian Language* (in Serbo-Croatian). Internal publication of the University of Novi Sad, 1972.

REFERENCES

Bradley, D. C. *Computational distinctions of vocabulary type*. Unpublished doctoral dissertation, MIT, 1978.

Carroll, J. B., & White, M. N. Age-of-acquisition norms for 200 picturable nouns. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 573-576. (a)

Carroll, J. B., & White, M. N. Word frequency and age of acquisition as determinants of picture-naming latency. *Quarterly Journal of Experimental Psychology*, 1973, *25*, 85-95. (b)

Chomsky, N. Remarks on nominalization. In R. A. Jacobs & P. S. Rosenbaum (Eds.), *Readings in English transformational grammar*. Waltham, Mass.: Ginn & Co., 1970.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335-359.

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance, VI*. London: Academic Press, 1977.

Forster, K. I. Accessing the mental lexicon. In E. C. T. Walker & R. J. Wales (Eds.), *New approaches to language mechanisms*. Amsterdam: North-Holland, 1976.

Forster, K. I. Levels of processing and the structure of the language processor. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing*. Hillsdale, N.J.: Erlbaum, 1979.

Forster, K. I., & Bednall, E. S. Terminating and exhaustive search in lexical access. *Memory & Cognition*, 1976, *4*, 53-61.

Glanzer, M., & Ehrenreich, S. L. Structure and search of the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 381-398.

Jaspersen, O. *Essentials of English grammar*. London: George Allen and Unwin, 1962.

Kostić, Dj. The structure of usage value of grammatical forms in Serbo-Croatian. *Report of the Institute of Experimental Phonetics and Speech Pathology*, Belgrade, 1965.

Labov, W. Gaining access to the dictionary. In J. F. Kavanagh & W. Strange (Eds.), *Speech and language in the laboratory, school and clinic*. Cambridge, Mass.: MIT Press, 1978.

Lukatela, G., Mandić, Z., Gligorijević, B., Kostić, A., Savić, M., & Turvey, M. T. Lexical decision for inflected nouns. *Language and Speech*, 1978, *21*, 166-173.

Lukatela, G., Savić, M., Gligorijević, B., Ognjenović, P., & Turvey, M. T. Bi-alphabetical lexical decision. *Language and Speech*, 1978, *21*, 142-165.

Lukatela, G., Savić, M., Ognjenović, P., & Turvey, M. T. On the relation between processing the Roman and Cyrillic alphabets: A preliminary analysis with bi-alphabetical readers. *Language and Speech*, 1978, *21*, 113-141.

Lukić, V. Active written vocabulary of pupils at the elementary school age (in Serbo-Croatian). Belgrade: Zavod za Izdavanje Udžbenika SR Srbije, 1970.

MacKay, D. G. Derivational rules and the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 1978, *17*, 61-71.

Manelis, L., & Tharp, D. The processing of affixed words. *Memory & Cognition*, 1977, *5*(6), 690-695.

Morton, J. Interaction of information in word recognition. *Psychological Review*, 1969, *76*, 165-178.

Morton, J. A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

Rubenstein, H. R., Lewis, S. S., & Rubenstein, M. A. Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 645-657.

Stanners, R. F., & Forbach, G. B. Analysis of letter strings in word recognition. *Journal of Experimental Psychology*, 1973, *98*, 31-35.

Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 399-412.

Stefanović, M. *Contemporary Serbo-Croatian language*. (in Serbo-Croatian). Belgrade: Naučna Knjiga, 1974.

Taft, M. Recognition of affixed words and the word frequency effect. *Memory & Cognition*, 1979, *7*, 263-272. (a)

Taft, M. Lexical access via an orthographic code: The basic orthographic syllable structure (BOSS). *Journal of Verbal Learning and Verbal Behavior*, 1979, *18*, 21-39. (b)

Taft, M., & Forster, K. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 638-647.

Whaley, C. P. Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 1978, *17*, 143-154.

A WORD SUPERIORITY EFFECT IN A PHONETICALLY PRECISE ORTHOGRAPHY

G. Lukatela,+ B. Lorenc,+ P. Ognjenović,+ and M. T. Turvey++

Abstract. Other things being equal, a letter is identified more accurately and rapidly in the context of a word than in the context of a nonword. This word-superiority effect has been demonstrated many times with materials conforming to English orthography. The present experiment, using the probe letter-recognition procedure, demonstrates the same effect for the Serbo-Croatian orthography. In that the English and Serbo-Croatian orthographies distinguish markedly in the level at which they systematically reference the spoken language, it appears that the word-superiority effect is not owing to orthographic idiosyncracies. Analysis of the effect in Serbo-Croatian suggests that it is not completely accountable for in terms of interletter probability structure and that word-specific factors may be involved.

Under the same conditions, a letter is identified more rapidly and more accurately in the context of a word than in the context of a nonword. This letter-in-context or word-superiority effect is now a well-established fact for fluent readers of the English orthography (Baron, 1978). Arguably, fluent readers of English relate more efficiently to English words than to letter strings with which they have had no experience because they have learned something about the structure of written English in general and/or the properties of English words in particular. What has been learned to enhance word perception cannot be precisely pinpointed. Nevertheless, several kinds of knowledge can be proposed as potential candidates, for example, meaning, whole-word familiarity, word-specific associations with sounds, spelling rules and familiarity with spelling patterns (Baron, 1978). Questions as to the aspect or aspects of word processing that these kinds of knowledge influence are largely unresolved, although most recent evidence appears to rule out the feature analysis of component letters (Krueger & Shapiro, 1979; Massaro, 1979; Staller & Lappin, 1979).

The major focus of the present paper is a simple question: Does the word superiority effect hold for an orthography that differs nontrivially from the orthography of English? Orthographies work as transcriptions of language because the patterning of symbols in written text bears a systematic relationship to some corresponding patterning in the spoken language. The orthography of English is principally (but not exclusively) systematic with reference to

---

+University of Belgrade.

++Also University of Connecticut.

Acknowledgment. This research was supported in part by NICHD Grant HD-08495 to the University of Belgrade and in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

the morphophonemics of the spoken language, while the orthography of Serbo-Croatian is principally (but not exclusively) systematic with reference to the (classically defined) phonemics of the spoken language (see Lukatela & Turvey, 1980; Lukatela, Popadić, Ognjenović, & Turvey, 1980). We might expect to find, therefore, differences between the reading-related processes exhibited by fluent readers of English and those exhibited by fluent readers of Serbo-Croatian. For fluent readers of Serbo-Croatian, lexical decision is mediated by phonetic recoding (Lukatela et al., 1980); in contrast, fluent readers of English tend to access the lexicon in nonphonological terms (Coltheart, Besner, Jonasson, & Davelaar, 1979). With respect to a distinction drawn by Baron and Strawson (1976), fluent readers of Serbo-Croatian may be disproportionately "Phoenician" (that is, treat the written word as an alphabetic transcription), while fluent readers of English may be disproportionately "Chinese" (that is, treat the written word as a logographic transcription). Though the latter contrast is exaggerated, it makes the point that the phonemically oriented Serbo-Croatian orthography and the morphophonemically oriented English orthography may give emphasis to different aspects of the written form of the word and thus motivate the acquisition of, and a dependency on, different kinds of knowledge for word perception. Perhaps the letter-in-context or word-superiority effect is indigenous to the English orthography (and to orthographies of like kind) and is due to the fact that the processing of written English often demands the use of recoding units larger than the single letter. We doubt that there is such a restriction on the word-superiority effect, but the question of the effect's dependency on the orthography must be asked nevertheless.

The question was addressed through the probe recognition procedure first introduced by Reicher (1969). A horizontally arranged string of letters is briefly exposed and followed immediately by a mask (covering the region of the letter string) together with two letters located above and below the position of a letter in the presented string. The subject's task is simply to choose which of the two letters occupied the probed position. Of interest is how letter recognition varies with the nature of the letter string.

Method

Subjects

The subjects were 41 undergraduate students from the Department of Psychology at the University of Belgrade who participated in the experiment as part of a course requirement. The majority of the subjects received their elementary education in eastern Yugoslavia, that is to say, they acquired the Cyrillic alphabet prior to the Roman alphabet (see Lukatela, Savić, Ognjenović, & Turvey, 1978).

Materials

The target letter strings and the response alternatives were Roman uppercase (see Lukatela et al., 1978), black letraset (Helvetia light, 12 pt) letters pressed onto the glass surface of 36-mm slides. Individual letters maximally subtended 21' x 25' of visual angle and the visual extent of a five letter string was 2°17' with the middle letter of the stimulus array positioned at the center of the display. The mask pattern subtended 21' vertical by 2°17' horizontal to coincide perfectly with the region occupied by the letter string. The response alternatives subtended 1°34' vertically from the top part of the upper letter to the bottom part of the lower letter. The light background regions of the target and mask fields were equated at 10 cd/m^2.

There were four kinds of target stimuli: single letters, five-letter words, five-letter nonwords with vowels ("pseudowords"), and five-letter nonwords without vowels ("nonwords"). Thirty-two instances of each kind were constructed. Six instances of each kind were used in the preliminaries to the experiment and twenty instances of each kind were used in the experiment proper.

In the fashion of Reicher (1969) and Wheeler (1970) the words and their response alternatives were selected so that the wrong alternative, if substituted for the probed letter, also made a word with a frequency of occurrence roughly equivalent to that of the target word. Frequency equivalence was determined according to the frequency count of Dj. Kostić (Note 1). Thus, if the target word were TAČKA (point), and the alternatives for the first letter as the probed letter were T and M, then the substitution of T by M would give MAČKA (cat).

The words were of five different consonant(c)-vowel(v) structures, CVCVC, CCVCV, VCCVC, VCVCV, CVCCV, which were represented in the set of twenty words, respectively, seven times, seven times, twice, twice, and twice. The different consonant-vowel structures were necessitated by the requirements that (1) only consonants were probed in the four kinds of stimuli (the nonwords were composed only of consonants) and (2) each letter position was probed equally often. Table 1 gives the words and pseudowords together with the response alternatives. Each of the twenty pseudowords was constructed from its word mate by changing two letters without altering the consonant-vowel structure. Which two letters were changed depended on the particular consonant-vowel structure of the word as is evident from inspection of Table 1. Moreover, the particular letter substitutes chosen were selected to keep the pronounceability of a word and its pseudoword partner approximately equivalent. This "pronounceability" stricture also determined the selection of the incorrect response alternative. The response alternatives for an individual pseudoword were the same as for its word mate.

The nonwords were constructed by a random drawing of consonants under the constraint that no letter could be repeated within a letter string. The single-letter stimuli were all consonants and they always occurred in the middle of the slide.

Procedure

A subject viewed sequences of slides presented by means of a three-channel tachistoscope (Scientific Prototype, Model GB) and responded to the critical member of a sequence by pressing one of two telegraph keys. The nearer of the two keys indexed "lower" and the farther of the two keys indexed "upper." A sequence of slides consisted of the following: Subsequent to a ready signal, a fixation field of 500 msec exposure was presented, followed by

Table 1

Words, pseudowords and response alternatives with target letters specified

WORDS	PSEUDOWORDS	RESPONSE ALTERNATIVES
HRANA	**H**REKA	H,G
LI**T**AR	LE**T**OR	T,M
S**R**EĆA	S**R**ISA	R,V
VRA**T**A	VLI**T**A	T,N
IZ**R**AZ	IG**R**EZ	R,L
NAPAD	**N**ALID	N,Z
U**L**ICA	U**L**EZA	L,D
TRA**V**A	TLE**V**A	V,K
SAVE**Z**	SAGI**Z**	Z,T
META**L**	MEBO**L**	L,K
O**B**RAZ	O**B**LEZ	B,D
GLAVA	**G**LOTA	G,S
BO**M**BA	BU**M**KA	M,R
KANA**L**	KASO**L**	L,P
PO**N**OĆ	PA**N**UĆ	N,M
O**P**ERA	O**P**INA	P,V
SVI**L**A	SRO**L**A	L,T
POJA**M**	PONE**M**	M,S
BRA**D**A	BLI**D**A	D,V
TAČKA	**T**AZLA	T,M

a slide containing one or five letters. The duration of this letter-string or target slide was tailored to the individual subject and therefore variable across subjects but constant for a given subject within the sequences of slides. Immediately following the termination of the target slide, that is, at an inter-stimulus interval of 0 msec, a slide containing a random patterning of lines (that overlapped the letters of the target slide) and two letters was presented for a duration of 1.5 sec. One of the two letters was above the masking pattern, while the other was below it. These two letters were aligned vertically and located so as to correspond to the position of one of the letters in the target slide. The subject's task was to press one of the two keys to identify which of the two letters, the upper or the lower, was the letter occurring in that position of the target slide. One of the letter alternatives was always correct.

The dependent measure was the accuracy of the subject's choice between the two response alternatives. A level of performance was sought, therefore, at which a subject recognized the probed-for letters above chance but not perfectly. To this purpose, the collection of data for analysis was preceded by a practice session during which the subject was familiarized with the task and during which the experimenter determined the duration of the target slide exposure at which the subject's performance was approximately seventy-five percent accurate.

The practice session was divided into two phases. During the first phase the exposure time of the target stimuli was held constant at 100 msec and the subject was given feedback on the accuracy of his or her choice. In the second phase the target stimulus duration was reduced until a duration yielding an accuracy of seventy-five percent was reached. Further sequences were then presented to assess the reliability of the criterial duration with increases or decreases introduced where necessary. Across subjects the duration yielding criterial performance ranged from 30 to 50 msec. Following the practice session forty sequences were presented to the subject with the target exposure at the individually determined duration and with the different types of stimuli distributed randomly.

Results and Discussion

The number of correct responses for each subject for each stimulus type was entered into a two-factor analysis of variance (Subject x Stimulus Type), which showed the type of stimulus to be significant, $F(3,123) = 12.69$, $p < .001$. The percentages of correct recognition for the four stimulus types were: single letters, 78.10; words, 81.19; pseudowords, 73.81; and nonwords, 64.52. Protected t-tests on the individual comparisons revealed a significant difference between words and nonwords ($p < .01$), words and pseudowords ($p < .02$), pseudowords and nonwords ($p < .01$) and single letters and nonwords ($p < .01$).

Let us consider first why we might *not* have expected a word-superiority effect for the Serbo-Croatian orthography. Suppose that the kind of knowledge that accounted for the effect in English was of the correspondence rules that parse script into the functional units to which phonemes can be systematically assigned. Venezky (1967, 1970) has given a detailed exposition of these rules for English. There are, of course, consistent mappings but they are often abstract and they generally relate graphic symbols to the morphophonemic and not to the phonetic level of the language. Moreover, their application generally involves lexical reference. Thus *sh* in *mishap* is not a single phoneme as it is in *ship* or *smash*. To know this the reader must recognize that in *mishap* the two letters are separated by a morpheme boundary. Knowledge of parts of speech in addition to morpheme identity is necessary for the pronunciation of *ate* at the end of words (compare the verbs *deflate*, *integrate* with the nouns *syndicate*, *frigate*). A more straightforward rule is that which ascribes the phoneme /s/ to c before e, i or y plus a consonant or juncture. Because of the opaqueness of English spelling it is often necessary for a speaker of English to communicate the spelling of a word that another finds perplexing by indicating precisely the identity and order of the alphabetic constituents. In contrast, a speaker of Serbo-Croatian can communicate the spelling in almost all cases by simply speaking the word more slowly. The point is that the fund of orthographic parsing rules required for spelling English has no equivalent in Serbo-Croatian and thus if such knowledge were a critical ingredient in the word-superiority effect, then no such effect should be expected in Serbo-Croatian.

Consider a further but related reason that derives from doubts as to the value of reforming the English orthography in the direction of greater phonetic specificity (cf. Gibson & Levin, 1975). Arguably, the efficient recognition of (English) words is principally based in the intra-word redundancies generated by orthographic rules. To increase the phonetic precision of a writing system is to strip away these important clues to a word's nature. The orthography of English allows skilled readers to obtain grammatical and semantic information about words from their orthographic forms (Chomsky, 1970). This is because English preserves the morphological similarity of words (for example, anxious, anxiety), whereas an orthography oriented to phonetics would forego, necessarily, this commitment to meaning and etymology. Thus in Serbo-Croatian even declensions of the same word may undergo orthographic modification in the interests of a phonetically precise transcription from the spoken to the written form (for example, noga, nozi, the nominative and dative forms, respectively, of the word meaning *leg*). Given these considerations one could entertain an argument of the following kind: Meaning is a type of knowledge that determines the word superiority effect. But meaning is less directly accessible from the internal structure of Serbo-Croatian words than it is from the internal structure of English words. At the time of making a choice in the probe recognition procedure, a reader of Serbo-Croatian is less likely to have accessed a letter string's meaning. Consequently, under the conditions of the task the meaning-based word/nonword distinction is less available to the Serbo-Croatian reader and thus the word-superiority effect less likely for the Serbo-Croatian orthography.

Of course, the arguments above are straw men. There is little if any reason for believing that the word-superiority effect is owing to a single factor operating in isolation so that the absence of that factor is sufficient to rule out the occurrence of the effect. Nevertheless, the arguments serve the purpose of underscoring differences between the two orthographies and what they entail in processing terms; the arguments suffice to indicate the kinds of rationalization that could be made if the perception of written Serbo-Croatian failed to manifest a superiority of words over nonwords. However, given that fluent readers of Serbo-Croatian did perceive letters in words better than letters in nonwords and pseudowords, let us proceed to consider the reasons why they did so. With regard to the nonsignificant difference between the words and the single letters, it suffices to note that when single letter performance is the poorer of the two (e.g., Carr, Lehmkuhle, Kottas, Astor-Stetson, & Arnold, 1976), it is probably due to positional uncertainty (Estes, 1975). In our experiment the single letters always occurred in the same position of the display.

That the words were perceived better than the nonwords may not require an appeal to word-specific factors in that the pseudowords were similarly superior. However, that the words were, in turn, perceived better than the pseudowords might mean that an appeal to word-specific factors may be required for a full account. The superiority in perception of words and pseudowords over nonwords can be considered from two perspectives: One emphasizes general orthographic distinctions and the other emphasizes general (non-orthographic) figural and conceptual distinctions between the two kinds of letter patterns. Thus the regularities of written Serbo-Croatian (for example, the tendency to alternate consonants and vowels, the limited number of consonant runs of two and three letters) present in the words and pseudowords and not present in the random consonant strings that were the nonwords may be the source of the perceptual distinction. Yet recourse to the regularities of the written language may be unnecessary; there are nonlinguistic factors that would distinguish the words and pseudowords from the nonwords in ways that are potentially exploitable by the perceiver.

Two categories of letters--vowels and consonants--comprised the words and pseudowords. One category of letters--consonants--comprised the nonwords and only one category of letters--consonants--was probed. There is much evidence to show that categorical information facilitates the detection of targets in visual search tasks (Brand, 1971; Ingling, 1972; Jonides & Gleitman, 1972, 1976; Lukatela et al., 1978). Sometimes referred to as a "conceptual" category effect, there is accumulating evidence that this may be an ill-chosen label. Denotable physical relations may well support the reliable discrimination of vowels from consonants (Staller & Lappin, 1979; White, 1977). At all events, the enhanced perception of letters in words and in pseudowords with respect to letters in nonwords may have been due to the ability to distinguish the target category (consonants) from the non-target category (vowels), thereby effectively reducing the number of letters to be processed. Staller and Lappin (1979, Experiment 4) provide one significant instance that this, indeed, can be the case.

Let us now consider the difference in perceptibility of words and pseudowords. The literature equivocates on the genuineness of word/pseudoword differences. There are a large number of studies reporting that both words and pseudowords are superior to nonwords but do not differ between themselves, and there are a large number of studies showing word/pseudoword differences (see Baron, 1978, for a review). The former suggest that the word superiority effect is due entirely to general properties of the structure of the written language that are manifest equally in words and pseudowords, while the latter suggest that factors specific to words do exist over and above the general properties common to words and pseudowords. Baron (1978) notes several possible reasons for this equivocality of which the following may speak to the present data. First, current knowledge does not permit a systematic equating of words and pseudowords on the many non-semantic, non-lexical dimensions of potential relevance to perceiving letter strings (for example, the frequencies of letter groups, the frequencies with which letter groupings occur in certain positions within the letter string). Second, methods vary in their sensitivity to the word-superiority effect and where the difference between words and nonwords is relatively small, that between words and pseudowords is usually nonexistent. Type of mask (Johnston & McClelland, 1973), visual angle of the display (Purcell, Stanovich, & Spector, 1978) and the onset asynchrony between letter string presentation and mask presentation (Michaels & Turvey, 1979) contribute significantly to the magnitude of the word-superiority effect.

The difference between words and pseudowords was significant in the present experiment. Is it a genuine word-specific effect? The answer is not easily given, largely because of the first reason noted above--ignorance of whether all the nonword-specific dimensions were equated between the two sets of stimuli. Nevertheless, when general factors are considered, such as frequency of letter patterns and geometric properties of the letter strings, there remains some reason for believing that specific factors such as meaning, lexical membership or whole-word familiarity (Baron, 1978) may have contributed to the word/pseudoword difference. With respect to geometric properties, Staller and Lappin (1979) have shown that the symmetry and directionality of letters are significant to the perceptibility of letters in letter contexts. In the present experiment, where a symmetrical letter (e.g., H,T) in a word was changed in the construction of its pseudoword pair, the letter was changed half of the time into another symmetrical letter and half of the time into a right-facing letter (e.g., G,L). Likewise, right-facing letters were converted into another right-facing letter half of the time and into a symmetrical letter the other half of the time. So at least in terms of these two dimensions, symmetry and directionality of individual letters, the words and pseudowords were numerically equated.

A potentially more significant and likely source of difference is the conditional probabilities among the letter pairs. Changing two letters of a word to produce a pseudoword may have changed the degree to which letter pairings conformed to the language. Consulting Tomić's (1978) digram frequency analysis of 1,250,000 tokens, the conditional frequencies of letter pairs in the forward direction (that is, the frequency the letter *b* occurs given letter i before it) were determined for each letter string. Since the strings were five letters in length, there were four conditional frequencies for each letter string; these four were summed for each individual string of letters. For the words of the present experiment the overall mean of the individual sums was 26,135 compared to an overall mean of 17,863 for the pseudowords. Moreover, of the twenty pairs of words and pseudowords, the word member was of higher summed conditional frequency in seventeen of the pairs. It would seem, therefore, that the word/pseudoword difference in the present experiment is accountable for in terms of differences in the interletter probability structure. A further analysis suggests, however, that interletter probability structure may not be the complete story.

A correlation computed between the summed conditional frequencies of pseudowords and the number of incorrect recognitions proved significant ($r = -.513$, $p < .05$), meaning that the higher the summed digram frequency the fewer the errors. In contrast, a similar correlation computed for the word stimuli proved insignificant ($r = -.005$). The possibility that interletter probability characteristics may have contributed more significantly to letter recognition in pseudowords than in words is consistent with other observations in the literature. Thus, Engel (1974) reported that the relationship between interletter probabilities and the accuracy of letter detection was most pronounced for low frequency words, and Rice and Robinson (1975) showed that the influence of mean digram frequency on lexical decision latencies was restricted to rare words. An analysis by Whaley (1978) concurs with these observations: Whereas general factors such as interletter probability structure contribute significantly to the perception of letter strings that are nonwords or pseudowords and perhaps to the perception of relatively new or unfamiliar words, they contribute relatively less significantly to the perception of words. In word perception the general aspects are overridden by the specific aspects such as richness of meaning and familiarity. In the absence of further analysis on general aspects we may, therefore, draw the qualified conclusions that the word-superiority effect of the present experiment is a word-specific effect.

It remains for us to make one final remark by way of reinforcing a point above with regard to the word/nonword data. The Serbo-Croatian language is biased heavily toward open syllables. A perusal of the Tomić (1978) norms reveals that consonant-vowel and vowel-consonant pairs are by far the most frequent, with consonant-consonant pairs comparatively rare. A crude comparison suggests that the relative proportion of consonant pairs and consonant triples in English is larger (Baddeley, Conrad, & Thompson, 1960, compared with Tomić, 1978). This difference between the interletter structure of the two languages may account for why the word/nonword difference in the present experiment was greater in magnitude than that generally reported for comparable experiments with English materials. In the present experiment with Serbo-Croatian the difference was roughly 17 percent compared to the difference commonly reported for English, which is on the order of 10 percent or less. Nonword letter strings composed solely of randomly selected consonants are considerably more like the internal structure of English words than they are like the internal structure of Serbo-Croatian words. Structurally speaking the difference between words and (all-consonant) nonwords is greater in Serbo-Croatian than it is in English.

To summarize, evidence has been provided for a word-superiority effect in the Serbo-Croatian orthography, an orthography that is markedly different from the English orthography in which the effect is most commonly reported. The Serbo-Croatian orthography is more closely related to (classical) phonemics, while the English orthography is more closely related to morphophonemics. The word-superiority effect, therefore, appears to be indifferent to the linguistic level referenced by the orthography. As with the word-superiority effect demonstrated in English (and see Dutch, 1980), the word-superiority effect demonstrated in Serbo-Croatian may resist explanation solely in terms of general properties of the written language.

REFERENCE NOTE

1. Kostić, Dj. *Frequency of occurrence of words in Serbo-Croatian.* Unpublished paper, Institute of Experimental Phonetics and Speech Pathology, Belgrade, Yugoslavia, 1965.

REFERENCES

Baddeley, A., Conrad, R., & Thompson, W. Letter structure of the English language. *Nature*, 1960, *186*, 414-416.

Baron, J. The word superiority effect. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes VI*. Hillsdale, NJ: Erlbaum, 1978.

Baron, J., & Strawson, C. Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 386-393.

Brand, J. Classification without identification in visual search. *Quarterly Journal of Experimental Psychology*, 1971, *23*, 178-186.

Carr, T. H., Lehmkuhle, S. W., Kottas, B., Astor-Stetson, E. C., & Arnold, D. Target position and practice in the identification of letters in varying contexts: A word superiority effect. *Perception & Psychophysics*, 1976, *19*, 412-416.

Chomsky, N. Phonology and reading. In H. Levin & J. P. Williams (Eds.), *Basic studies on reading*. New York: Basic Books, 1970.

Coltheart, M., Besner, D., Jonasson, J. T., & Davelaar, E. Phonological encoding in the lexical decision task. *Quarterly Journal of Experimental Psychology*, 1979, *31*, 489-508.

Dutch, S. *Predicting letter search time through words and nonwords: The role of statistical frequency and lexical status in the word-superiority effect*. Unpublished doctoral dissertation, University of Connecticut, 1980.

Engel, G. R. On the functional relationship between word identification and letter probability. *Canadian Journal of Psychology*, 1974, *28*, 300-309.

Estes, W. K. The locus of inferential and perceptual processes in letter identification. *Journal of Experimental Psychology: General*, 1975, *105*, 122-145.

Gibson, E. J., & Levin, H. *The psychology of reading*. Cambridge, Mass.: The MIT Press, 1975.

Ingling, N. Categorization: A mechanism for rapid information processing. *Journal of Experimental Psychology*, 1972, *94*, 239-243.

Johnston, J. C., & McClelland, J. L. Visual factors in word recognition. *Perception & Psychophysics*, 1973, *14*, 365-370.

Jonides, J., & Gleitman, H. A conceptual category effect in visual search: O as a letter or a digit. *Perception & Psychophysics*, 1972, *12*, 457-460.

Jonides, J., & Gleitman, H. The benefit of categorization in visual search. Target location without identification. *Perception & Psychophysics*, 1976, *20*, 289-298.

Krueger, L. E., & Shapiro, R. G. Letter detection with rapid serial visual presentation: Evidence against word superiority at feature extraction. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 657-673.

Lukatela, G., Popadić, D., Ognjenović, P., & Turvey, M. T. Lexical decision in a phonologically shallow orthography. *Memory & Cognition*, 1980, *8*, 124-132.

Lukatela, G., Savić, M., Ognjenović, P., & Turvey, M. T. On the relation between processing the Roman and the Cyrillic alphabets: A preliminary analysis with bi-alphabetical readers. *Language and Speech*, 1978, *21*, 113-141.

Lukatela, G., & Turvey, M. T. Some experiments on the Roman and Cyrillic alphabets of Serbo-Croatian. In J. Kavanaugh & R. Venezky (Eds.), *Orthography, reading and dyslexia*. Baltimore, Md.: University Park Press, 1980.

Massaro, D. W. Letter information and orthographic context in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 595-609.

Michaels, C. F., & Turvey, M. T. Central sources of visual masking: Indexing structures supporting seeing at a single, brief glance. *Psychological Research*, 1979, *41*, 1-61.

Purcell, D. G., Stanovich, K. E., & Spector, A. Visual angle and the word superiority effect. *Memory & Cognition* 1978, *6*, 3-8.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, *81*, 275-280.

Staller, J. D., & Lappin, J. S. Word and nonword superiority effects in a letter detection task. *Perception & Psychophysics*, 1979, *25*, 47-54.

Tomić, T. Statistička analiza srpskohrvatskog teksta pomoću računara [statistical analysis of Serbo-Croatian text by computer]. Institut za jezike i književnost u Sarajevu, Odeljenje za jezike, Posebna izdanja, 1978, *4*, 221-235, Sarajevo.

Venezky, R. English orthography: Its graphical structure and its relation to sound. *Reading Research Quarterly*, 1967, *2*, 75-105.

Venezky, R. *The structure of English orthography*. The Hague: Mouton, 1970.

Whaley, C. P. Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 1978, *17*, 143-154.

Wheeler, D. D. Processes in word recognition. *Cognitive Psychology*, 1970, *1*, 59-85.

White, M. J. Identification and categorization in visual search. *Memory & Cognition*, 1977, *5*, 648-657.

LARYNGEAL ACTIVITY IN ICELANDIC OBSTRUENT PRODUCTION*

Anders Löfqvist+ and Hirohide Yoshioka++

Abstract. Laryngeal activity in the production of voiceless obstruents and obstruent clusters in Icelandic was investigated by the combined techniques of transillumination and fiberoptic filming of the larynx. Contrasts of preaspirated, unaspirated, and postaspirated voiceless stops were found to be produced basically by differences in laryngeal-oral timing. During clusters of voiceless obstruents, one or more continuous laryngeal opening and closing gestures occurred depending on the segments in the cluster. Peak velocity of glottal abduction was higher for fricatives than for stops. This, and other differences in laryngeal adjustments and interarticulator timing between stops and fricatives are most likely due to different aerodynamic requirements for stop and fricative production. The present results further question the usefulness of timeless feature descriptions for modeling speech production.

INTRODUCTION

The present study deals with two topics in speech production that will be discussed from two different perspectives. The first topic is laryngeal activity in speech, in particular the organization of laryngeal abduction and adduction in voiceless obstruent production. Production of voiceless obstruents requires not only certain laryngeal adjustments but also the formation of a closure or constriction in the vocal tract that is made by adjusting supralaryngeal articulators. Since obstruent production thus involves simultaneous activity at both laryngeal and supralaryngeal levels, the laryngeal and oral articulations have to be coordinated in time. The second topic to be dealt with is laryngeal-oral coordination in obstruent production.

Following the title of this Conference we will discuss these two topics from a Nordic and a general perspective. The Nordic perspective is that of the phonetics of Icelandic. Icelandic is, in a sense, a rich language since it has contrasts of preaspirated, unaspirated and postaspirated voiceless stops. We will thus discuss laryngeal activity and interarticulator programming in Icelandic, and examine how they are used to produce the acoustic signals that are required by the phonology of the language.

---

*A version of this paper was presented at the Fourth International Conference of Nordic and General Linguistics, Oslo, 23-27 June, 1980, and will appear in Nordic Journal of Linguistics.

+Also Lund University, Lund, Sweden.

++Also University of Tokyo, Tokyo, Japan.

Acknowledgment: This work was supported by NINCDS Grants NS-13617 and NS-13870, and by BRS Grant RR-05596.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

We will also discuss these problems from a more general point of view, trying to extract some general properties of laryngeal function in speech that appear to be used by speakers of different, and unrelated languages. If such universal aspects of laryngeal behavior in speech can be found, they are likely to reflect general properties of the organization of the speech motor system.

Finally, we will address the general problem of interarticulator programming in speech. If we loosely define speech as audible movements, it behooves us to account for temporal and spatial aspects of their coordination and control. We will thus argue that speech production should be viewed as an instance of control of coordinated movements in general, and outline what we think is a powerful and productive theoretical approach to this problem.

The aim of the present study is thus twofold: To contribute to a better understanding of laryngeal control and interarticulator programming in Icelandic, and to use the Icelandic data to evaluate and develop further current models of laryngeal and motor behavior in speech.

METHOD

Procedure

Laryngeal adjustments were monitored simultaneously by fiberoptic filming and transillumination. Filming was made through a flexible fiberscope (Olympus VF Type O) at a film speed of 60 frames/second. The fiberscope, inserted through the nose, was kept in position by a specially designed headband. A synchronization signal was recorded on one channel of a multichannel instrumentation tape recorder for frame identification. Relevant portions of the film were analyzed frame by frame with a computer assisted analyzing system, and the distance between the vocal processes was measured as an index of glottal opening.

The light from the fiberscope was used as part of a transillumination system, whereby the amount of light passing through the glottis was sensed by a phototransistor (Philips, BPX 81) placed on the surface of the neck just below the cricoid cartilage, and held in position by a neckband. The signal from the transistor was amplified and recorded on one channel of the tape recorder.

The transillumination signal was processed with the Haskins Laboratory system (Kewley-Port, 1977). The signal was rectified, integrated over a 5 msec interval, and sampled at a rate of 200 Hz for further computer processing. For averaging, the signal was aligned with reference to a predetermined, acoustically defined line-up point.

In order to calculate the speed of glottal opening change, the signal was smoothed over a 15 msec interval. The velocity was then calculated by successive subtractions at 5 msec increments.

The measurements from the film were compared with the transillumination signals obtained for the same tokens of the test utterances. No further processing was applied to the measurements from the film.

A direction-sensitive microphone was used to record the audio signal in direct mode on one channel of the instrumentation recorder. The audio signal was sampled at 10 kHz and used for determination of the line-up points as well as for acoustic measurements. This signal was then rectified and analyzed in parallel with the biomechanical signals. In the averaging process the rectified audio signal was integrated over 15 msecs.

Linguistic Material

The linguistic material consisted of Icelandic voiceless obstruents and obstruent clusters, with a word boundary preceding, following or intervening within the cluster. Both the transillumination technique and fiberoptic filming require a wide pharyngeal cavity, which had to be taken into account in selecting the linguistic material. Icelandic words were used, and these words are given in Table 1. The words in Set A were placed in the frame "Segðu ..." ("Say..."). All the words in Set B were combined with those in Set C and placed in the carrier "En ..." ("But ...") to yield 24 normal Icelandic sentences.

Table 1

The linguistic material. The words in set A contain contrasts of preaspirated (left column), unaspirated (middle column), and postaspirated (right column) voiceless stops. All the words in set B were combined with those in set C to provide different obstruent clusters.

Set A

seppi	biti	penni
hitti	dimmi	tunnu

Set B

Elli
Rut
Agnes
mest ... Ágúst
dóttir Eiríks
sonur prests

Set C

ýtir
sýtir
kítir
spýtir

A native female speaker from Southern Iceland read the material 12 times from randomized lists. Five to twelve repetitions of each utterance type were used for averaging. Fiberoptic films were made during 3 to 6 of these repetitions.

RESULTS

Figure 1 compares the patterns of glottal opening obtained by transillumination and by fiberoptic filming of four utterances. A good agreement between the two methods is apparent. This was also shown by a correlation analysis. For each of 95 utterances, a Pearson product moment correlation coefficient was calculated between the two curves. The correlation coefficients were highly significant ($0.6<r<0.7$ for 4 utterances; $0.7<r<0.8$ for 10 utterances; $0.8<r<0.9$ for 29 utterances; $r>0.9$ for 52 utterances, with $P<0.001$ in all cases).

Figure 2 presents averaged transillumination signals and audio envelopes for three different types of voiceless stops, unaspirated, postaspirated, and preaspirated. They differ in at least two dimensions of laryngeal activity. First, the relative timing of glottal abduction/adduction and oral closure/release is different. For the unaspirated stop, glottal abduction starts at the implosion, and peak glottal opening, i.e., glottal adduction, occurs close to the implosion. The postaspirated type has glottal abduction beginning at implosion and peak glottal opening at the oral release. For the preaspirated stop, both glottal abduction and peak glottal opening precede oral closure.

A second difference illustrated in Figure 2 is that of glottal opening size. Although the amplitude information of the transillumination signal should be interpreted with great caution due to technical problems, the present data suggest that voiceless postaspirated stops have larger glottal opening than their preaspirated and unaspirated cognates. Glottal opening is smaller for the preaspirated type, and very small for the unaspirated one. For the latter, the fiberoptic films revealed a small, spindle-shaped opening in the membraneous portion of the glottis. Figure 2 also indicates an even larger glottal opening for the voiceless fricative in "seppi."

Average transillumination and acoustic records of consonant clusters are shown in Figure 3. The average records in Figure 3 only contain tokens with similar cluster duration, and where no pause signaled the location of the word boundary. In other cases, the cluster durations showed large variability, as will be discussed further below.

One feature of the clusters in Figure 3 is that laryngeal adjustments can be organized in one or more continuous opening and closing gestures. When only one gesture occurs, its timing relative to supralaryngeal events varies depending on the segments involved. In clusters of stop + fricative, or fricative + stop ("Elli spýtir," "Rut sýtir," "mest ýtir"), peak glottal opening occurs during the fricative. 'Long' fricatives as in "Agnes sýtir," and "Eiríks spýtir" also have one glottal gesture.

More than one laryngeal gesture occurs in clusters of fricative + aspirated stop, or fricative + stop + fricative (e.g., "Agnes kítir," "mest sýtir," "mest spýtir"). In these cases, the timing of laryngeal and oral articulations is similar to that found in single stops or fricatives, i.e., peak glottal opening occurs close to onset of the fricatives and close to release for aspirated stops.

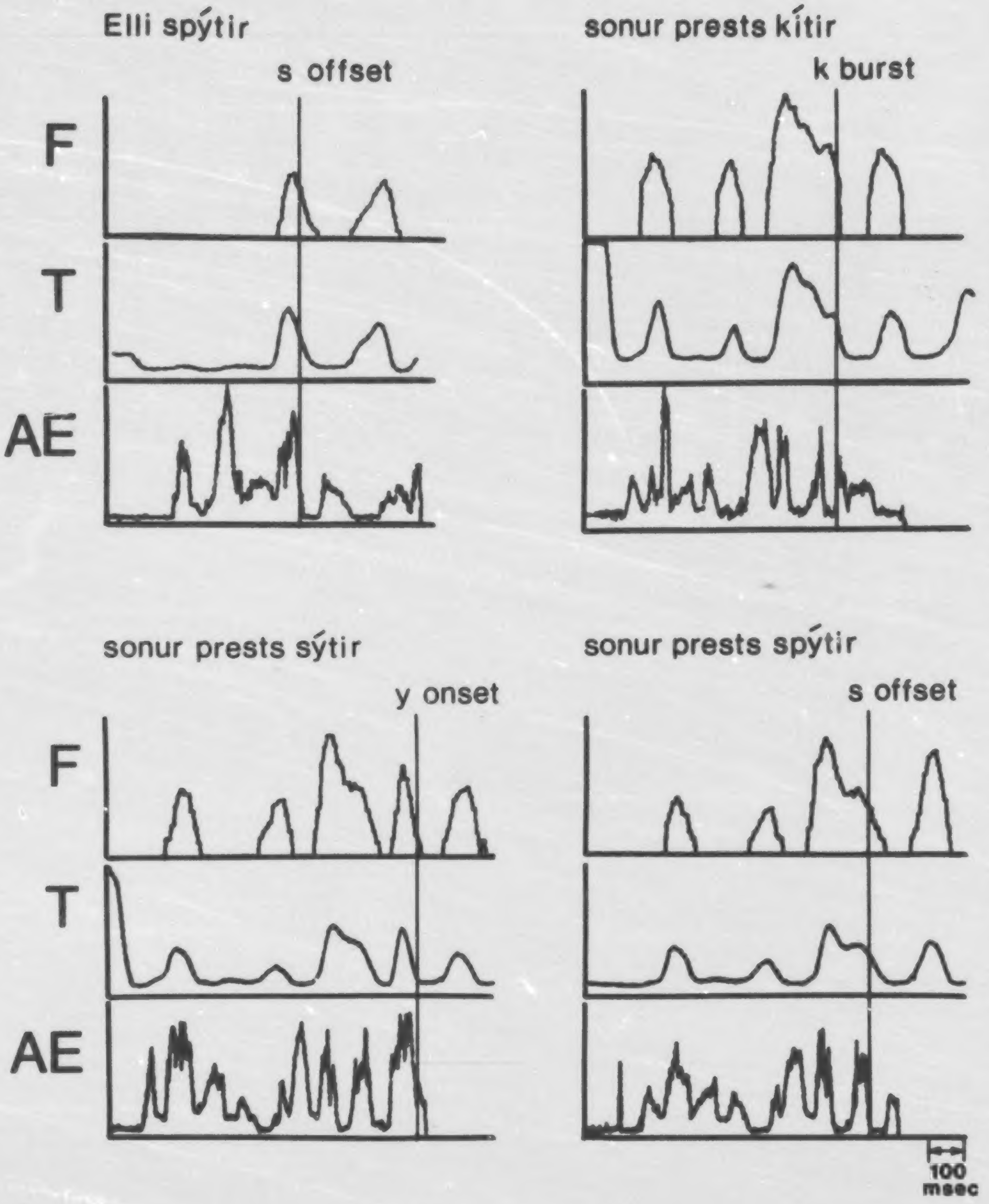


Figure 1. Comparisons of fiberoptic and transillumination records for four utterances. F = glottal area obtained by fiberoptic filming. T = glottal area obtained by transillumination. AE = audio envelope.

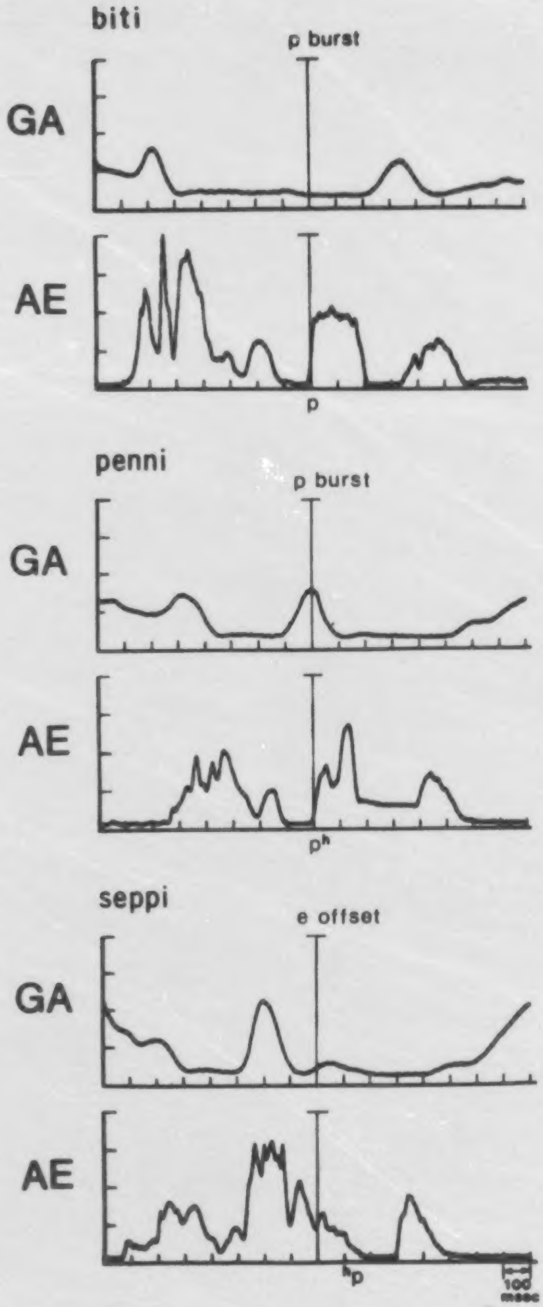


Figure 2. Average transillumination signal (GA), and audio envelope (AE) for utterances containing unaspirated (top), postaspirated (middle), and preaspirated (bottom) stops.

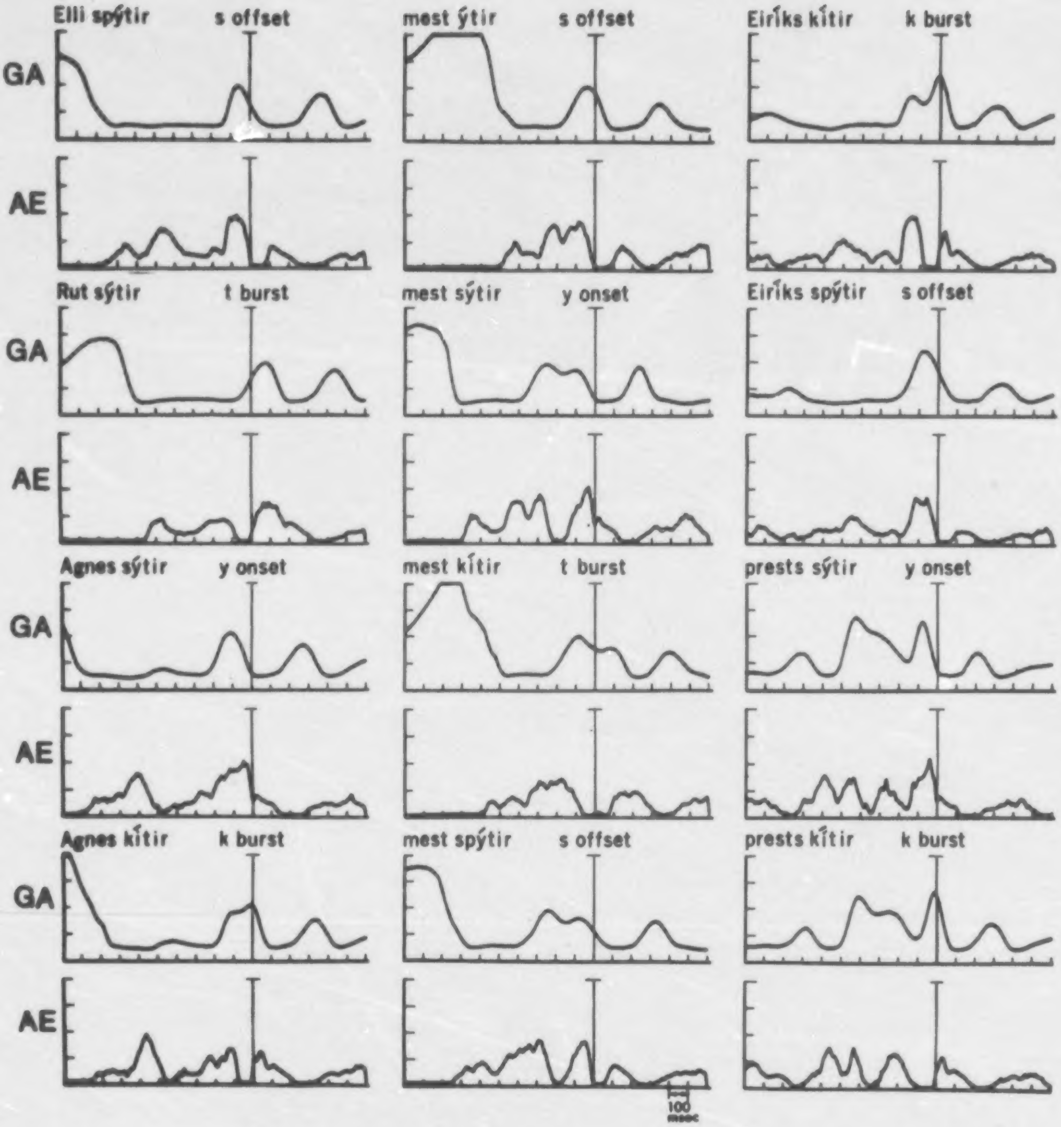


Figure 3. Glottal area and audio signals for 12 utterances containing different obstruent clusters.

As mentioned above, some cluster durations showed rather large variability between tokens. This is illustrated further in Figures 4 and 5, which show single tokens of two utterance types. In both cases a unimodal pattern is changed into a bimodal one as the duration of the cluster increases. For the longest durations of "Agnes spýtir" a silent pause intervened between the two words. In these cases the glottis was completely adducted, whereas in all other cases where more than one opening gesture occurred, the glottis was only slightly adducted without complete closure between the two opening maxima.

A closer view of glottal opening and velocity is presented in Figures 6, 7, and 8 for selected single obstruents and obstruent clusters. The displacement averages were made with an integration time of 15 msecs, and all the curves are aligned with reference to the offset of the preceding vowel. In the velocity plots, positive values indicate abduction and negative values indicate adduction.

The word initial vowels in the test material were generally produced with a glottal attack. In Figures 7 and 8, utterances containing a glottal attack following the obstruents are shown with solid lines, and a tight glottal closure for the attack is evident in the displacement plots.

Figure 6 shows some clear differences between stops and fricatives. A comparison between the utterance containing a word initial stop ("kítir") and those with a word initial fricative ("sýtir," and "spýtir") shows that for the stop, peak glottal opening occurs later than for the fricative. Similarly, peak velocity of the abduction gesture occurs closer to vowel offset for the fricative than for the stop. Peak abduction velocity is also higher for the fricatives.

Similar differences between clusters beginning with stops and fricatives are shown in Figure 7 ("Rut ...," and "Agnes ..."). Peak glottal opening and peak velocity of the abduction gesture occur closer to vowel offset for the clusters beginning with a fricative, and these clusters also show higher velocity of the abduction gesture. For the clusters beginning with a fricative, the abduction velocity shows a single, narrow peak, whereas for the clusters beginning with a stop the peak is broader.

These differences between clusters beginning with stops and fricatives are less clear in Figure 8, as far as timing of peak glottal opening and peak abduction velocity are concerned. This is presumably related to the very short closure duration for /k/ in "Eiríks," where a closure was absent even in the acoustic record for some tokens.

A further observation in Figures 7 and 8 is also of interest. For a given set of utterances within a graph, peak velocity of the abduction gesture occurs more or less at the same time with respect to offset of the preceding vowel. This holds true irrespective of variations in speed, size, duration, and timing of the glottal gesture.

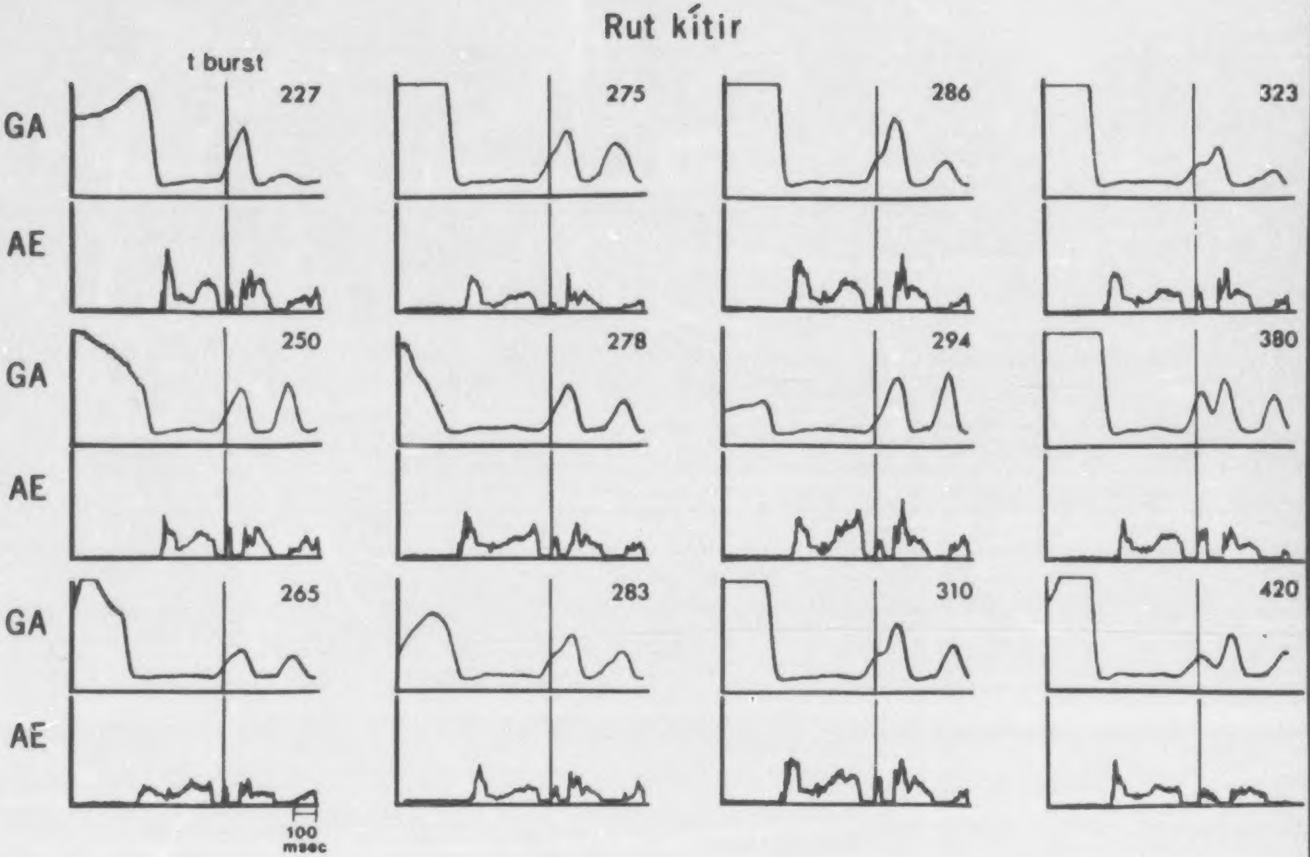


Figure 4. Glottal area and audio signals for 12 tokens of the utterance "En Rut kítir." Numbers at right in each graph indicate duration (in milliseconds) of the cluster /t#k/.

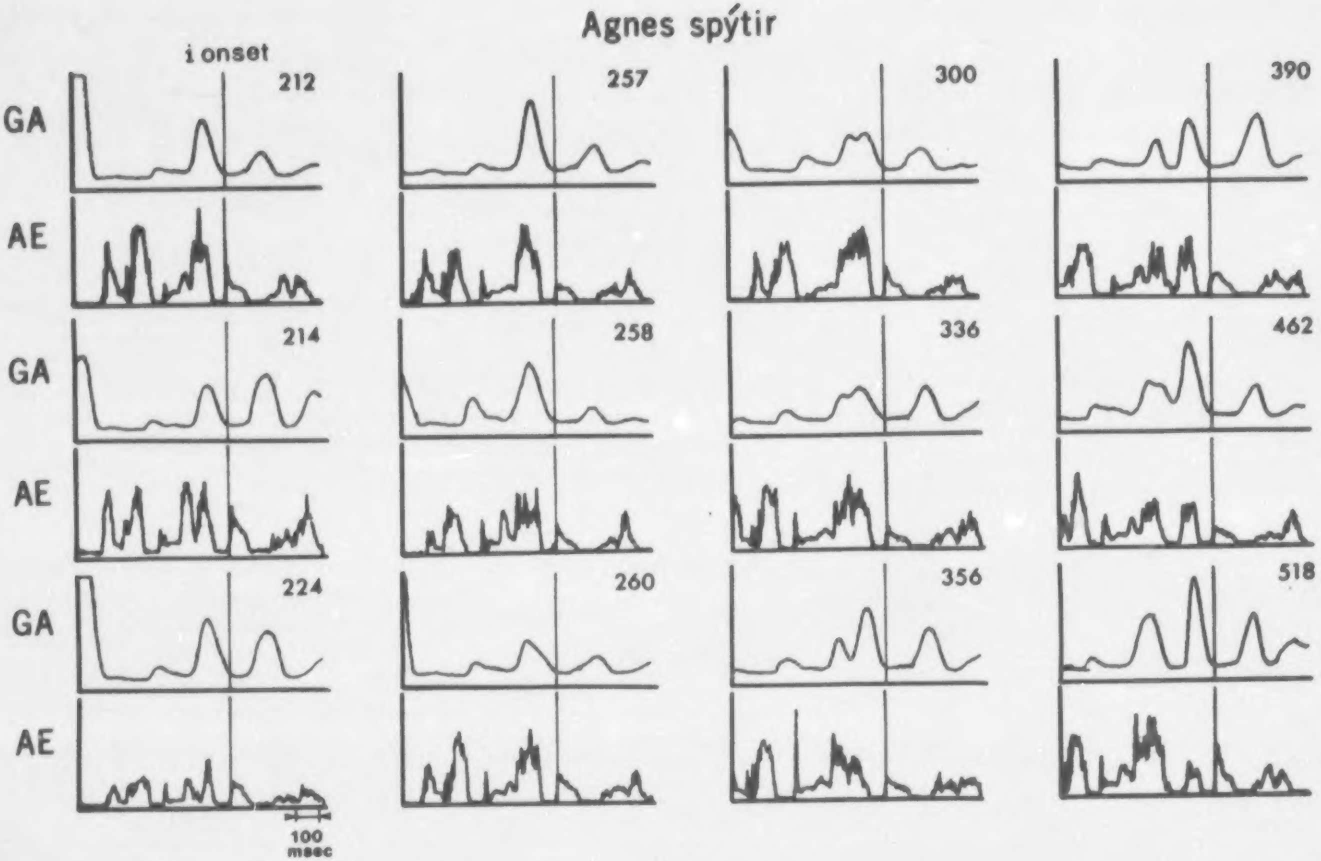


Figure 5. Glottal area and audio signals for 12 tokens of the utterance "En Agnes spýtir." Numbers at right in each graph indicate duration (in milliseconds) of the cluster /s#sp/.

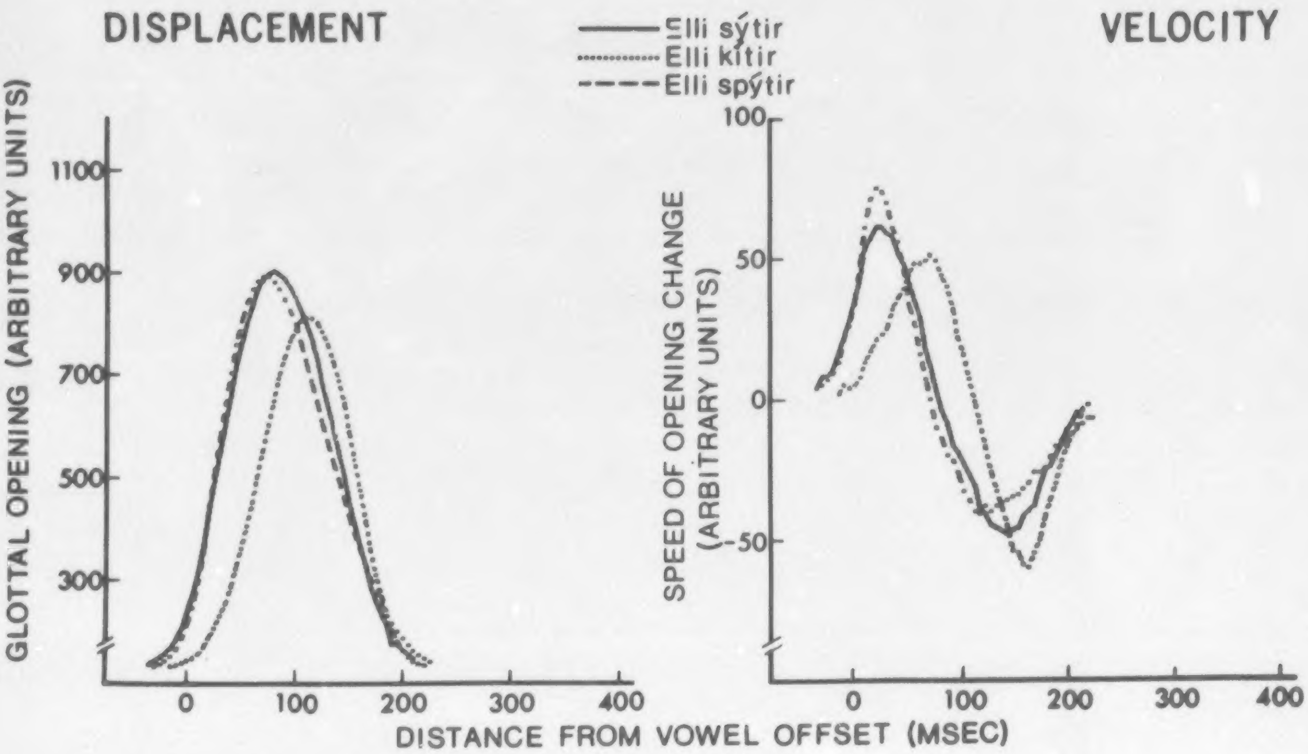


Figure 6. Plots of size and speed of the glottal abduction/adduction gesture for three obstruents. Zero on x-axis indicates offset of the vowel preceding the obstruents. Abduction velocity is shown with positive sign, adduction velocity with negative sign.

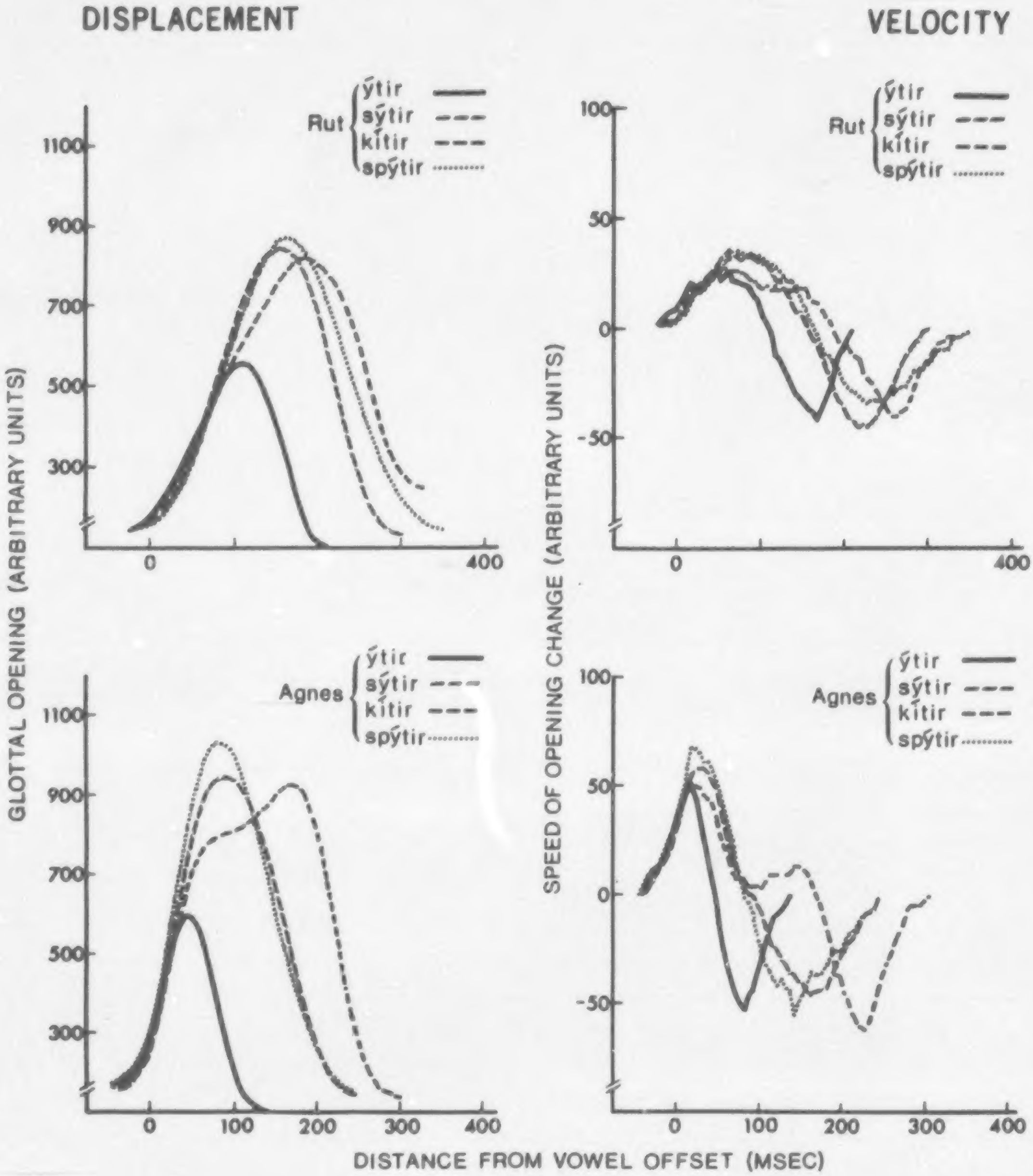


Figure 7. Plots of size and speed of the glottal abduction/adduction gesture for eight different obstruents and obstruent clusters. Zero on x-axis indicates offset of the vowel preceding the obstruents. Abduction velocity is shown with positive sign, adduction velocity with negative sign.

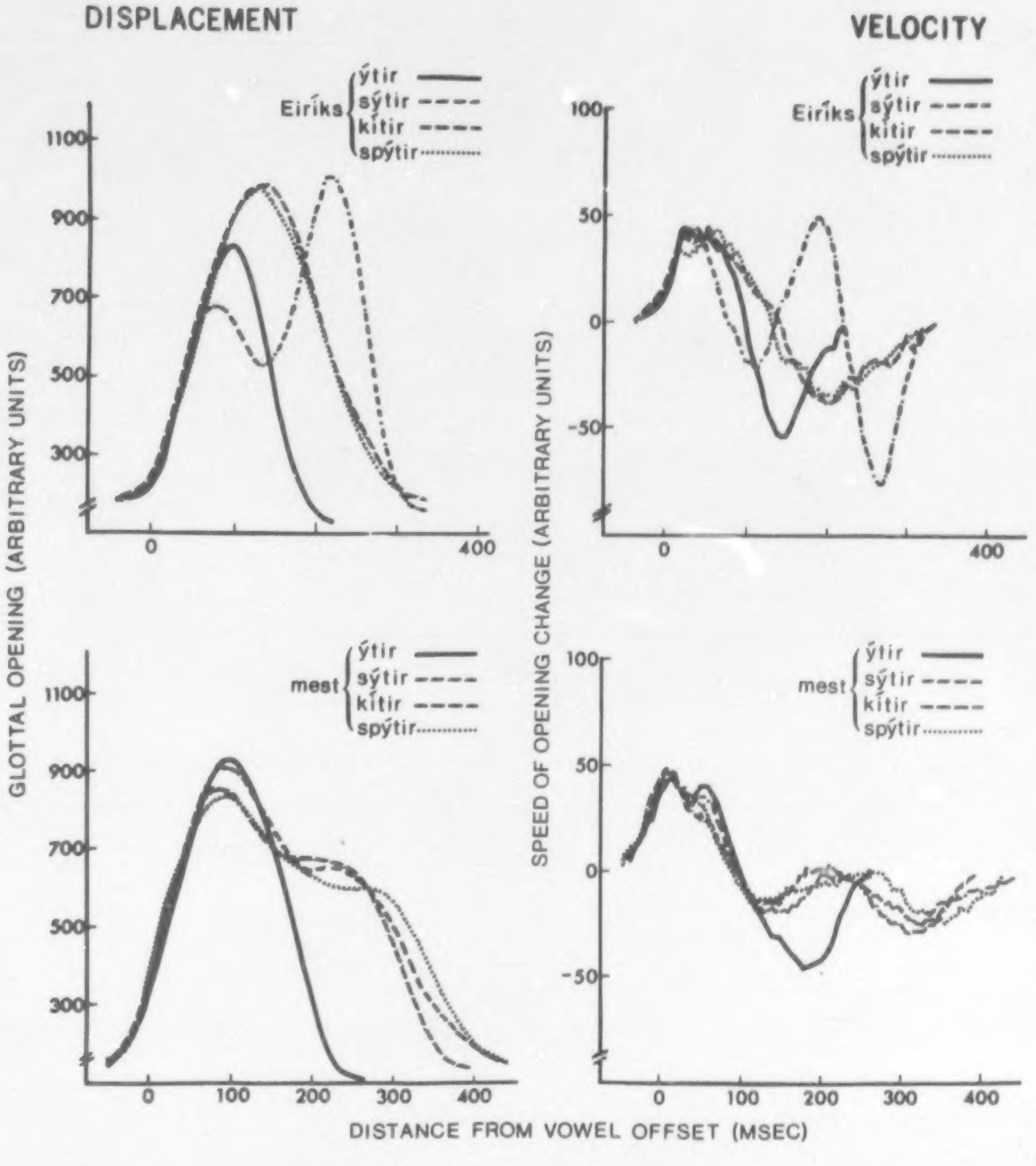


Figure 8. Plots of size and speed of the glottal abduction/adduction gesture for eight different obstruent clusters. Zero on x-axis indicates offset of the vowel preceding the obstruents. Abduction velocity is shown with positive sign, adduction velocity with negative sign.

DISCUSSION

The present results are limited to a single subject, and may thus contain speaker specific elements. They are, however, in good agreement with those obtained from another Icelandic speaker by Pétursson (1976, 1978). Moreover, they also agree with other cross-language data, and would thus seem to show some general aspects of laryngeal behavior in speech.

Concerning the phonetics of Icelandic, the differences in laryngeal activity between preaspirated, unaspirated and postaspirated stops are similar to those presented by Pétursson (1976). In one respect, the present material would seem to show some speaker specific traits in that peak glottal opening occurs close to, or coincides with, stop release in postaspirated stops. For the subject investigated by Pétursson (1976), peak glottal opening precedes stop release by a longer interval for the same stops. This variation is also reflected in longer VOT values for this stop category in the present study, about 80 milliseconds compared to 40-50 milliseconds in Pétursson's study. Such interspeaker variability should come as no surprise, given the variability permitted by the linguistic code. Since similar acoustic signals can be produced using different articulatory strategies, this may be another source of interspeaker variation. The exact timing of peak glottal opening relative to oral release in postaspirated stops would seem to differ between languages depending on the amount of aspiration required by the phonology of the language, and also between speakers, since different combinations of interarticulator timing and glottal aperture size can result in similar durations of aspiration.

As for the production of voiceless obstruent clusters, the present material further validates the conclusions, based on American English and Swedish material (Yoshioka, Löfqvist, & Hirose, 1979; Löfqvist & Yoshioka, 1980) on the organization of laryngeal activity in speech. During a voiceless cluster, when the glottis is open for a long period, variations in glottal opening occur. Laryngeal articulation is thus organized in one or more continuously changing opening and closing gestures. The general rule governing the occurrence of one or more gestures seems to be that sounds requiring a high rate of air flow and/or buildup of oral pressure are produced with a separate gesture. To judge from the results of the American English and Swedish studies, these gestures are actively controlled by muscular adjustments, and are not passive results of aerodynamic forces.

From Figures 4 and 5, it appears that a word boundary marked by a silent pause is associated with glottal adduction. It is possible that such an adduction is made to prevent air flow and waste of air during an ongoing utterance. Another interpretation would be that word boundaries are in themselves accompanied by glottal adduction. A 'long' fricative spanning a word boundary can, however, be produced with one or two gestures, cf. Figure 5. Glottal adduction is thus not necessarily associated with linguistic boundaries. Adduction is also found in certain clusters without apparent boundaries, where it seems better ascribed to segmental properties.

We would favor a unified account of laryngeal activity that reflects both the organization of the speech motor system and the encoding of linguistic information. Static glottal open configurations rarely seem to occur in speech, and also appear difficult to maintain in some nonspeech conditions (cf. Löfqvist, Baer, & Yoshioka, 1980). A continuously changing glottis thus seems to be a basic feature of laryngeal control. The laryngeal gestures are precisely coordinated with supralaryngeal events to meet the aerodynamic requirements for producing a signal with a specified acoustic structure.

Before we turn to a discussion of the displacement and velocity data presented in Figures 6, 7 and 8, it is appropriate to discuss briefly the acoustic consequences of differences in interarticulator timing at implosion and explosion of voiceless obstruents.

Glottal abduction in voiceless obstruents contributes to cessation of glottal vibrations and, by reducing laryngeal resistance to air flow, to the high air flow and/or buildup of oral pressure. In voiceless stops, initiation of the abduction before oral closure produces preaspiration as shown in Figure 2. If glottal abduction starts after oral closure, prevoicing results, and if the abduction gesture occurs after stop release, a voiced (or murmured) aspirated stop is produced. Similarly, different timing relationships between glottal adduction and oral release produce contrasts of unaspirated and postaspirated stops. These different contrasts of aspiration and voicing are thus basically produced by differences in interarticulator timing. At the same time, differences in size of glottal aperture, similar to those shown in Figure 2 between unaspirated and postaspirated voiceless stops, often co-occur with the timing differences.

In Figures 6 and 7 we noted certain differences between stops and fricatives in the displacement and velocity patterns of the laryngeal adjustments. In particular, peak glottal opening occurs closer to offset of the preceding vowel and the opening velocity is higher for the fricative. Another difference is also evident, i.e., glottal abduction starts later relative to the offset of the preceding vowel for the stop. Some of these differences are most likely related to aerodynamic requirements for stop and fricative production. A rapid increase in glottal area would allow for the high air flow necessary to generate the turbulent noise source during voiceless fricatives (Stevens, 1971). In stops, a slower increase in glottal opening together with the concomitant oral closure could be sufficient to stop glottal vibrations and allow the buildup of oral pressure. As noted above, the timing of glottal opening during stop closure is part of the mechanism controlling aspiration (cf. Löfqvist, in press).

The present results are less clear for the velocity of the adduction gesture. There is a tendency for the closing speed to be higher when peak glottal opening occurs close to the onset of the following vowel. Closing speed is also rather high before a glottal attack.

Peak velocity of the abduction gesture tends to occur at more or less the same point in relation to vowel offset for stops and fricatives, respectively, irrespective of variations in velocity, size and duration of the gesture. Similar constant relationships between offset of a preceding vowel and the occurrence of peak velocity of glottal abduction have been found in Japanese (Yoshioka, Löfqvist, & Hirose, 1980) and also in American English and Swedish. This would indicate that the beginning of the initial acceleration of glottal abduction is the same.

The present results provide further illustration of a tight temporal coordination of laryngeal and oral articulations in voiceless obstruent production. The nature of this coordination constitutes an important problem for any theory of speech production.

Models of speech production based on feature spreading (Daniloff & Hammarberg, 1973; Hammarberg, 1976; Bladon, 1979; see also Fowler, 1980) would seem incapable of handling this kind of interarticulator programming, at least in their current form. One reason is that their temporal resolution is limited to quanta of phone or syllable size, whereas laryngeal-oral coordination in obstruents requires a finer grain of analysis. An additional problem is that it is unclear how such models can be interfaced with a theory of control of coordinated movements, since they do not specifically address the general problem of interarticulator coordination in space and time. These limitations of feature spreading models stem partly from the fact that they take as input the units of linguistic analysis. Linguistic feature descriptions usually lack an intrasegmental temporal domain, whereas the present results indicate that such a domain is necessary, at least for some classes of speech sounds.

As interarticulator timing appears to be an essential feature of voiceless obtruent production, one may question the descriptive adequacy of feature systems with timeless representations for modeling speech production, whatever their merits may be for abstract phonological analysis. Specifying glottal states along dimensions of spread/constricted glottis and stiff/slack vocal cords (Halle & Stevens, 1971) would thus seem not only to be at variance with the phonetic facts but also to introduce unnecessary complications. The difference between aspirated and unaspirated stops is one of timing rather than of spread versus constricted glottis. Similarly, the difference between voiceless and voiced aspirated stops is also one of timing rather than of stiff versus slack vocal cords. Preaspirated stops are naturally accounted for within a timing framework but cannot be readily differentiated from postaspirated ones in a timeless feature representation. Even though the size and speed of the glottal abduction and adduction gesture is a controlled variable, this gesture does not occur randomly in obstruent production but is tightly coordinated with supraglottal events. The importance of interarticulator timing in obstruent production is not a new idea, e.g., Rothenberg (1968), Lisker and Abramson (1971), Ladefoged (1973), and it has also been noted by phonologists favoring timeless phonological descriptions (e.g., Anderson, 1974).

Given the dynamic character of speech production and the need to coordinate different articulators in space and time, a theory of speech production should account for both these aspects. One view of motor control that incorporates these features is the theory proposed by Bernstein (1967) and elaborated by Greene (1971, 1972; see also Boylls, 1975; Turvey, 1977; Kugler, Kelso, & Turvey, 1980; Kelso, Holt, Kugler, & Turvey, 1980; Fowler, Rubin, Remez, & Turvey, 1980). Designed to cope with the number of degrees of freedom to be directly controlled, this theory views motor coordination in terms of constraints between muscles or groups of muscles that have been set up for the execution of specified movements. Areas of motor control where this theory has proved to be productive include locomotion (Grillner, 1975), posture control (Nashner, 1977), and hand coordination (Kelso, Southard, &

Goodman, 1979). One merit of this view is that it predicts and rationalizes tight temporal relationships between articulators. In particular, it predicts that some such relationships should remain invariant across changes in stress and speaking rate, and material presented by Tuller and Harris (1980) on oral articulators is in agreement with this prediction. One aspect of the present results would seem to fit into this theoretical framework. Peak velocity of the glottal abduction gesture was found to occur almost at the same point in time relative to the offset of a preceding vowel. It is conceivable that this fixed temporal relationship is a feature of the control of laryngeal-oral coordination. Under this interpretation, we would expect similar fixed relations between aspects of supralaryngeal articulatory movements and the laryngeal gestures. Work in progress will further clarify this issue.

REFERENCES

Anderson, S. R. *The organization of phonology*. New York: Academic Press, 1974.

Bernstein, N. *The co-ordination and regulation of movements*. New York: Pergamon, 1967.

Bladon, R. A. Motor control of coarticulation: Linguistic considerations. In *Proceedings of the Ninth International Congress of Phonetic Sciences*, Vol. 2, 1979, 325-331.

Boylls, C. C. A theory of cerebellar function with applications to locomotion. II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat. *COINS Technical Report 76-1*, (Department of Computer and Information Science, University of Massachusetts), 1975.

Daniloff, R., & Hammarberg, R. On defining coarticulation. *Journal of Phonetics*, 1973, *1*, 239-248.

Fowler, C. A. Coarticulation and theories of extrinsic timing. *Journal of Phonetics*, 1980, *8*, 113-133.

Fowler, C. A., Rubin, P., Remez, R., & Turvey, M. Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production*. New York: Academic Press, 1980, 373-420.

Greene, P. H. Introduction. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. L. Tsetlin (Eds.), *Models of the structural-functional organization of certain biological systems*. Cambridge, Mass.: MIT Press, 1971, xi-xxxi.

Greene, P. H. Problems of organization of motor systems. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 2). New York: Academic Press, 1972, 303-338.

Grillner, S. Locomotion in vertebrates: Central mechanisms and reflex interaction. *Physiological Reviews*, 1975, *55*, 247-304.

Halle, M., & Stevens, K. A note on laryngeal features. *Quarterly Progress Report* (Research Laboratory of Electronics, Massachusetts Institute of Technology), 1971, *101*, 198-213.

Hammarberg, R. The metaphysics of coarticulation. *Journal of Phonetics*, 1976, *4*, 353-363.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland Publishing Company, 1980, 49-70.

Kelso, J. A. S., Southard, D. L., & Goodman, D. On the coordination of two-handed movements. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 229-238

Kewley-Port, D. EMG signal processing for speech research. *Haskins Laboratories Status Report on Speech Research*, 1977, *SR-50*, 123-146.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland Publishing Company, 1980, 3-47.

Ladefoged, P. The features of the larynx. *Journal of Phonetics*, 1973, *1*, 73-83.

Lisker, L., & Abramson, A. Distinctive features and laryngeal control. *Language*, 1971, *47*, 767-785.

Löfqvist, A. Interarticulator programming in stop production. *Journal of Phonetics*, in press.

Löfqvist, A., Baer, T., & Yoshioka, H. Scaling of glottal opening. *Journal of the Acoustical Society of America*, 1980, *67*, S53. (Abstract)

Löfqvist, A., & Yoshioka, H. Laryngeal activity in Swedish obstruent clusters. *Journal of the Acoustical Society of America*, 1980, *68*, 792-801.

Nashner, L. M. Fixed patterns of rapid postural responses among leg muscles during stance. *Experimental Brain Research*, 1977, *30*, 13-24.

Pétursson, M. Aspiration et activité glottale. *Phonetica*, 1976, *33*, 169-198.

Pétursson, M. Jointure au niveau glottale. *Phonetica*, 1978, *35*, 65-85.

Rothenberg. M. *The breath-stream dynamics of simple-released-plosive production*. (Bibliotheca Phonetica, Vol. 6). Basel: Karger, 1968.

Stevens, K. Airflow and turbulence noise for fricative and stop consonants: Static considerations. *Journal of the Acoustical Society of America*, 1971, *50*, 1180-1192.

Tuller, B., & Harris, K. S. Temporal models of interarticulator programming. *Journal of the Acoustical Society of America*, 1980, *67*, S64. (Abstract)

Turvey, M. T. Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting and knowing*. Hillsdale, N.J.: Erlbaum, 1977, 211-265.

Yoshioka, H., Löfqvist, A., & Hirose, H. Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-59/60*, 127-151.

Yoshioka, H., Löfqvist, A., & Hirose, H. Laryngeal adjustments in Japanese voiceless sound production. *Journal of the Acoustical Society of America*, 1980, *67*, S52. (Abstract)

LARYNGEAL ADJUSTMENTS IN JAPANESE VOICELESS SOUND PRODUCTION*

Hirohide Yoshioka+, Anders Löfqvist++ and Hajime Hirose+++

Abstract. As part of a series of investigations on the production of sequences of unvoiced sounds in different languages, the current experiment was conducted using the combined techniques of photoelectric glottography, fiberoptic filming and laryngeal electromyography. Particular attention was paid to devoiced vowel production in various voiceless consonantal environments including geminates. The data show that the glottal opening gesture during a voiceless sequence containing a devoiced vowel is characterized by a uni-modal pattern, unless the vowel occurs between a voiceless fricative and a geminated one, as in /siQs/, where a bimodal pattern may occur. The movement results also suggest that the velocity and size of the glottal opening gesture vary according to the nature of the adjacent voiceless obstruents: The speed of the opening phase is slow when a stop precedes the vowel, and fast when a fricative precedes it. The peak glottal opening attained during the devoiced vowel is larger when a fricative either precedes or follows than when the vowel is surrounded on both sides by single or geminated stops. Furthermore, it is revealed that the peak velocity of the initial opening gesture occurs at almost the same time in relation to the voicing offset of the preceding vowel, regardless of the properties of the surrounding voiceless obstruents and, thus, irrespective of variations in the magnitude of velocity and opening size.

INTRODUCTION

At the 97th Meeting of the Acoustical Society of America, we reported how voiceless sound sequences, such as voiceless obstruent clusters, are organized in terms of their glottal opening and closing gestures, using native speakers of American English (Yoshioka, Löfqvist, & Hirose, 1979) and Swedish (Löfqvist & Yoshioka, 1980). The conclusion of those studies was that, in the production of sequential unvoiced sounds, the glottal opening gesture is characterized by a one, two, or more-than-two-peaked pattern in a regular fashion according to the nature of the voiceless segments: A voiceless obstruent specified by aspiration or frication noise tends to require a

---

*A version of this paper was presented at the 99th Meeting of the Acoustical Society of America, Atlanta, Georgia, 21-25 April 1980.

+Also University of Tokyo, Tokyo, Japan.

++Also Lund University, Lund, Sweden.

+++University of Tokyo, Tokyo, Japan.

Acknowledgment: This work was supported by NINCD Grants NS-13617 and NS-13870, and BRS Grant RR-05596 to Haskins Laboratories.

separate opening gesture, while an unaspirated stop in a voiceless environment can be produced within the opening gesture attributed to an adjacent aspirated stop or fricative. For example, an /sk#sk/ sequence in English was produced in most cases with two separate opening gestures. In contrast, an /sks#k/ string was in general accompanied by three opening gestures (Yoshioka et al., 1979).

Furthermore, the velocity of the initial opening movement was shown to vary depending on the properties of the initial voiceless segment: When the first unvoiced segment in the cluster was a fricative, the speed of the opening movement was significantly faster than when the initial voiceless sound was an aspirated or unaspirated stop, regardless of the nature of the following voiceless segments. This also meant that the difference in velocity during the initial abduction phase held true despite the fact that, for most clusters beginning with a voiceless unaspirated stop, peak glottal opening occurred during a following fricative segment.

In order to examine the validity of these notions across different languages, the current experiment was carried out using the same combined techniques of photo-electric glottography, fiberoptic filming and laryngeal electromyography, in cooperation with a native speaker of Japanese. The phonology of Japanese does not allow voiceless "pure" obstruent clusters other than geminates. Syllable-final obstruents also rarely occur in this language. On the other hand, in conversational speech of the Tokyo dialect there is a well-known phenomenon of vowel devoicing in that a high vowel, such as /i/ and /u/, surrounded by voiceless obstruents on both sides is often produced without any vocal fold vibrations during the vowel segment (e.g., Hattori, 1951; Han, 1962; Fujimura, 1971; Sawashima, 1973). Therefore, we paid particular attention to devoiced vowel production in various voiceless consonantal environments including geminates.

METHOD AND PROCEDURE

The techniques used in the present experiment were simultaneous recordings of photo-electric glottography, fiberoptic filming and laryngeal electromyography (EMG), in parallel with the audio signal.

The EMG data were obtained using bipolar hooked-wire electrode techniques (Basmajian & Stecko, 1962; Hirano & Ohala, 1969). The electrodes, consisting of a pair of platinum-tungsten alloy wires (50 microns in diameter with isonel coating), were inserted perorally into the posterior cricoarytenoid muscle (PCA) under indirect laryngoscopy with the aid of a specially designed curved probe (Hirose, Gay, Strome, & Sawashima, 1971). Before insertion, topical anesthetic was applied to the mucous membrane of the hypopharynx using a small amount of 4% Lidocaine spray (Xylocaine). For verification of electrode position, the subject was instructed to perform several non-speech and speech maneuvers that are well understood in terms of PCA involvement, such as inspiration and expiration, swallowing, pitch changes including register shifts, glottal attacks, and voiced-voiceless sound contrasts. The EMG signal was monitored on an oscilloscope not only during the verification gestures but also during the entire recording session.

The interference voltages of the EMG signals, after high-pass filtering at 80 Hz, were recorded on a multichannel FM recorder together with the audio signal. After full-wave rectification and integration over a 5-msec time window, the action potentials were fed into a computer at a sampling rate of 200 Hz for further processing to obtain the muscle activity patterns for ensemble-averaged tokens with a 35-msec time constant (Kewley-Port, 1977). The figures to be presented in this paper represent activity patterns aligned with reference to the voicing offset of the vowel preceding the voiceless sequence.

For the movement data, the glottal view through a flexible laryngeal fiberscope (Olympus VF-O type, 4.5 mm in outer diameter) was photographed with a cine camera at a rate of 60 frames/sec. A synchronization signal was registered on the FM recorder to identify each frame. Then, frame by frame analyses were made with the aid of a mini-computer to calculate the distance between the vocal processes; this distance is considered one of the indicators of glottal width (Sawashima & Hirose, 1968; Sawashima, 1976).

A cold DC light source (Olympus CLS), providing illumination of the upper glottal area, also served as the light source for the photo-electric glottography. The amount of light passing through the glottis was sensed by a phototransistor (Philips BPX 81) placed on the neck just below the lower edge of the cricoid cartilage. The electrical output was recorded on another channel of the FM tape. These signals were sampled at 200 Hz and processed on the computer.

A native male speaker of the Tokyo dialect, one of the authors, served as the subject. Among the various voiceless environments surrounding a devoiceable vowel /i/, the combination of /s/ and /k/ is optimum in forming the greatest possible number of meaningful words in Japanese. Therefore, as is shown in Table 1, we chose the test words that contain a devoiceable vowel /i/ in the middle of voiceless obstruents composed of the phonemes /s/ and /k/. For example, the production of the first word in this list, /kikee/, which means "anomaly" in Japanese, may be transcribed as having an unvoiced string [ki̥k] -- a [k] plus a devoiced [i̥] plus a slightly aspirated [k].

For the first 2-3 repetitions of each test word, embedded in the frame sentence "sorewa ------desu," "we call it ------," simultaneous recordings of EMG, photo-electric output and fiberoptic filming were made together with the audio signals, followed by 14-28 additional recordings of only EMG and photo-electric signals. During the latter part of the session, the glottal image was constantly monitored through the fiberoptic viewfinder. Such careful monitoring is mandatory to obtain reliable interpretations of large amounts of photo-electric recordings, as we have discussed elsewhere (Yoshioka et al., 1979; Löfqvist & Yoshioka, 1980).

RESULTS

Figure 1 illustrates the results for the test word /siQsee/. Since the glottal opening patterns obtained by photo-electric glottography have been shown to be practically identical to those obtained by plotting the distance between the vocal processes from the fiberoptic cine-films, we will focus on

Table 1

Test words and the carrier sentence (Q = geminate phoneme).

「それは＿＿です」 “sorewa____desu” 〔unvoiced string〕

奇　形	/kikee/	〔ki̥k〕
吉　景	/kiQkee/	〔ki̥kk〕
規　制	/kisee/	〔ki̥s〕
吉　世	/kiQsee/	〔ki̥ss〕
詩　形	/sikee/	〔ʃi̥k〕
失　敬	/siQkee/	〔ʃi̥kk〕
姿　勢	/sisee/	〔ʃi̥s〕
失　政	/siQsee/	〔ʃi̥ss〕

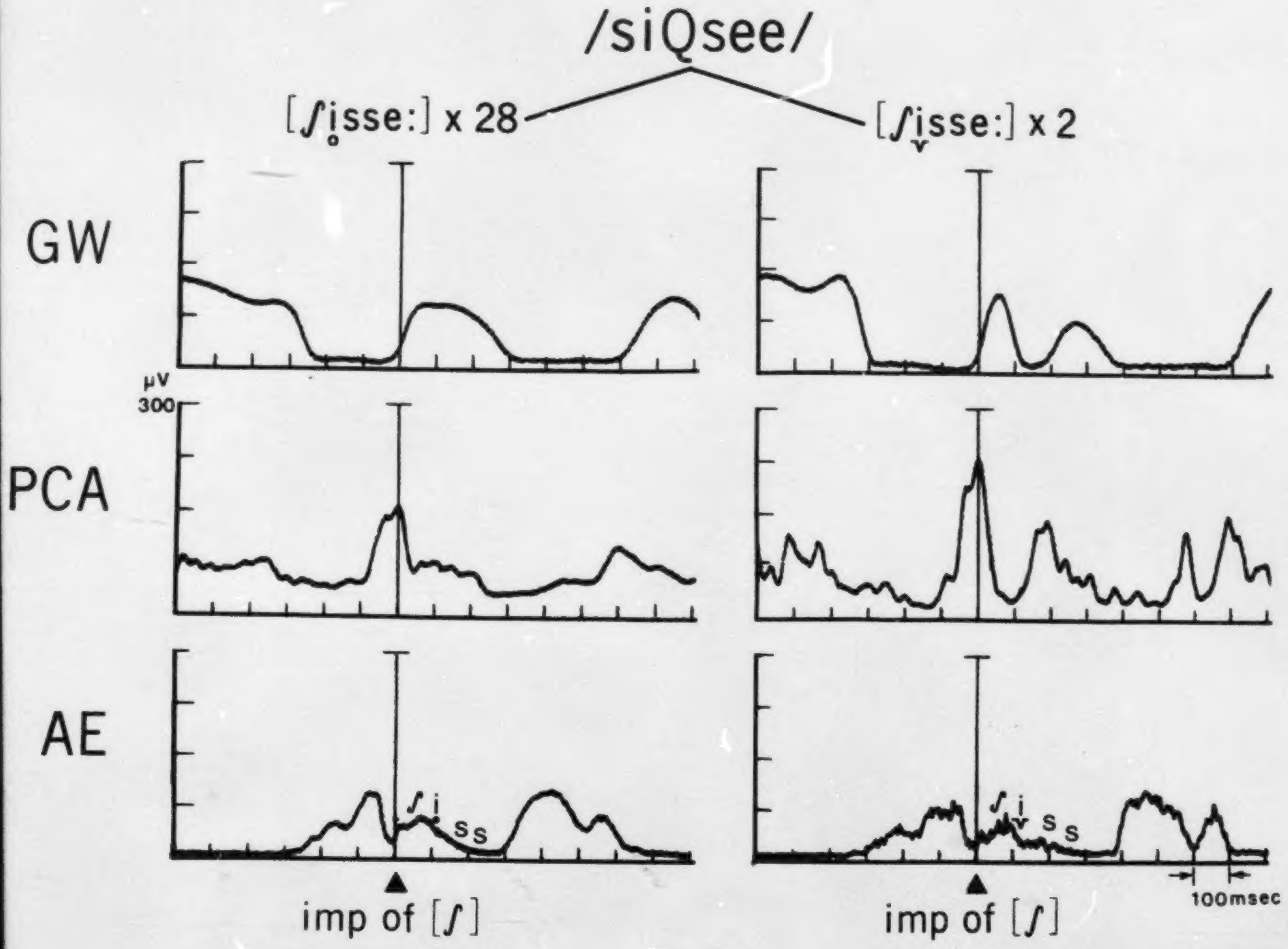


Figure 1. Averaged glottograms, PCA activity patterns and audio envelope curves for the test sentence containing the test word /siQsee/. 28 devoiced tokens (left), and 2 voiced tokens (right), respectively.

the photo-electric glottograms as an index of glottal width change during the pertinent voiceless sequence productions. Here, the top row (GW) represents the averaged glottograms for two allophonic groups: One is devoiced, and the other is voiced. Among the 30 repetitions, 28 tokens were produced with devoiced vowels, while only two of them had fully voiced vowels. These variations were easily detectable in audio waveforms, in sound spectrograms and by listening to the recorded tape. As for laryngeal EMG, the Figure contains the corresponding averaged activity patterns of the abductor muscle--the posterior cricoarytenoid muscle (PCA)--that has been demonstrated to substantially control glottal aperture (Hirose, 1976; Yoshioka, 1979). These signals were aligned with respect to the voicing offset of the preceding vowel. It is obvious that, when it is uttered with a fully voiced vowel, two clear separate glottal opening gestures are found for the /siQs/ production at both the movement and the electromyographic levels. In contrast, the averaged curves for the devoiced group is a little unclear. The abductor muscle (PCA) activity curve in the middle may be characterized by two opening gestures: The first is associated with a high and steep peak, followed by a second that is broad but of moderate activity level. The glottographic pattern for this devoiced group at the top is more complicated, in that one might describe it as having two peaks or, alternatively, a sort of plateau.

Since all the other test words, except the one containing /siQs/ mentioned above, were always produced with a devoiced vowel, all the averaged curves henceforth from Figure 2 are those for completely devoiced groups. Figure 2 shows the averaged glottographic pattern and the corresponding averaged abductor muscle activity pattern for the devoiced /sis/ sequence in comparison with those for the devoiced /siQs/ shown in Figure 1 and repeated in Figure 2. Here, several points are worth mentioning. First, the averaged glottogram for the non-geminated /sis/ is clearly distinguished by a uni-modal curve, while that for the geminated /siQs/ is characterized by a broad or bimodal pattern as mentioned above. This finding seems to be reflected in the EMG: The averaged PCA activity curve for the non-geminated /sis/ has a single peak around the line-up point, while that for the geminated /siQs/ is, as mentioned before, characterized by two separate activity patterns. In addition, despite the differences in the overall modality between these two utterance types at both movement and EMG levels, the initial opening phases are quite similar: The peak glottal openings are approximately of the same size and are reached almost at the same time. As for the PCA activity, both the curves have their peaks around the same time, i.e., the line-up point.

Figure 3 shows the activity patterns for a devoiced vowel /i/ surrounded on both sides by a pair of single or geminated stops. In comparison with those for the devoiced vowel /i/ occurring between voiceless fricatives shown in Figure 2, the glottographic curves for these cases have a single, smaller peak. The slopes of the glottographic curves for this stop group are also more gradual than for those surrounded by voiceless fricatives in Figure 2. This shows that the glottal opening gesture during the voiceless sequence containing a devoiced vowel may vary according to the nature of the surrounding consonants; slow for the voiceless stop and fast for the voiceless fricative.

Figures 4 and 5 show the patterns for the utterance types that contain a devoiceable vowel /i/ in a voiceless sequence composed of two different

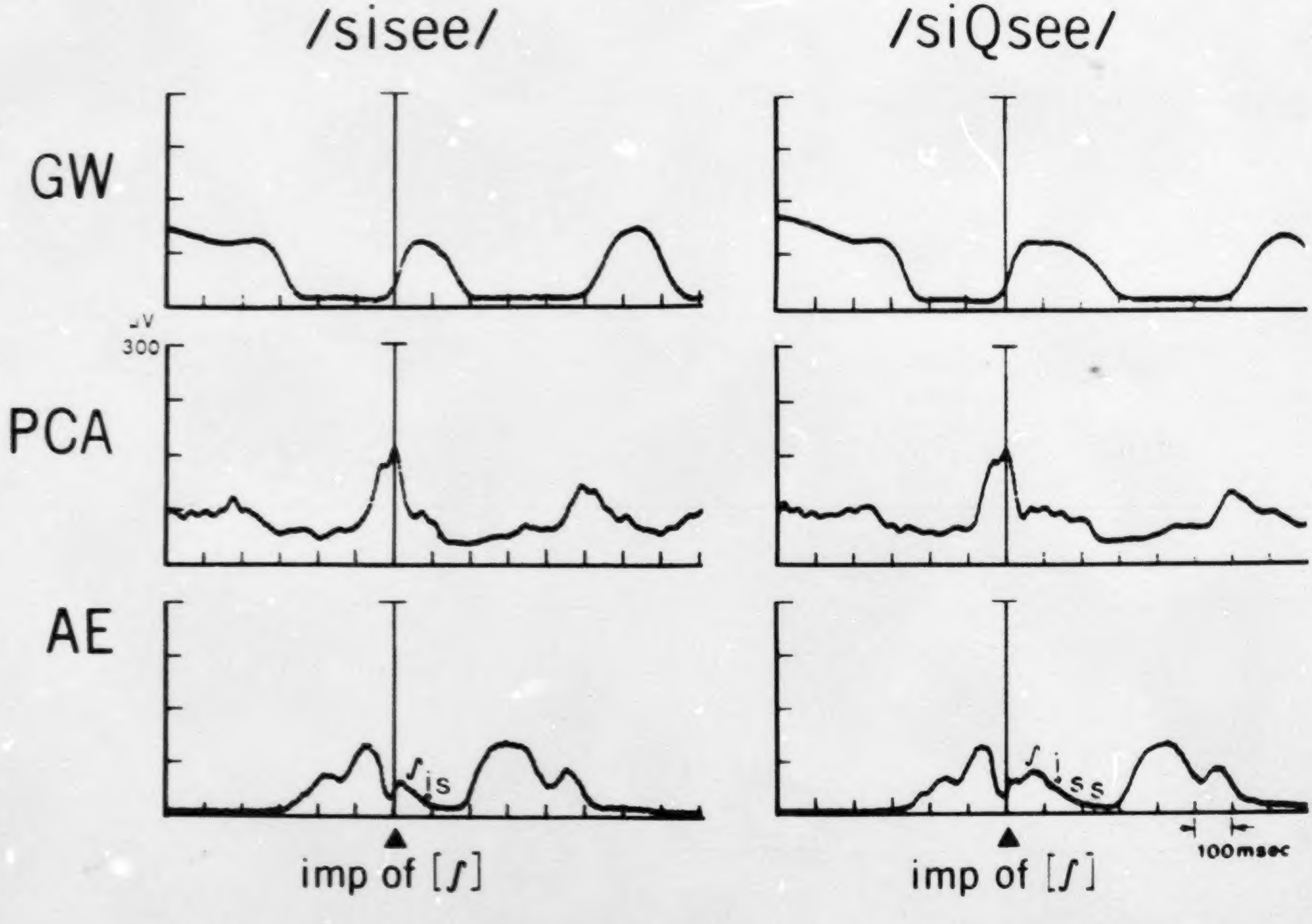


Figure 2. Averaged glotttograms, PCA activity patterns and audio envelope curves for the sentences containing the devoiced test words /sisee/ and /siQsee/, respectively.

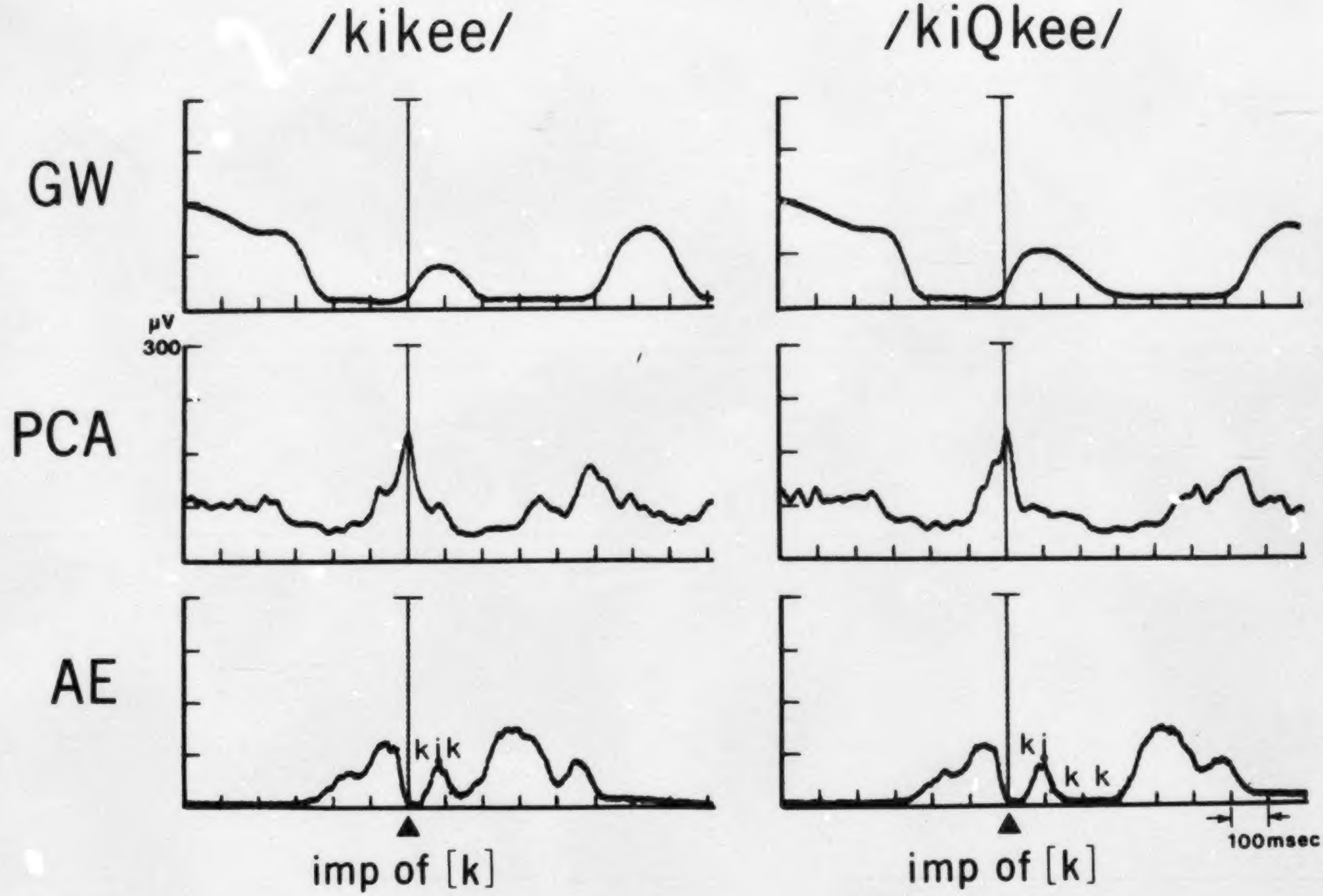


Figure 3. Averaged glottograms, PCA activity patterns and audio envelope curves for the sentences containing the devoiced test words /kikee/ and /kiQkee/, respectively.

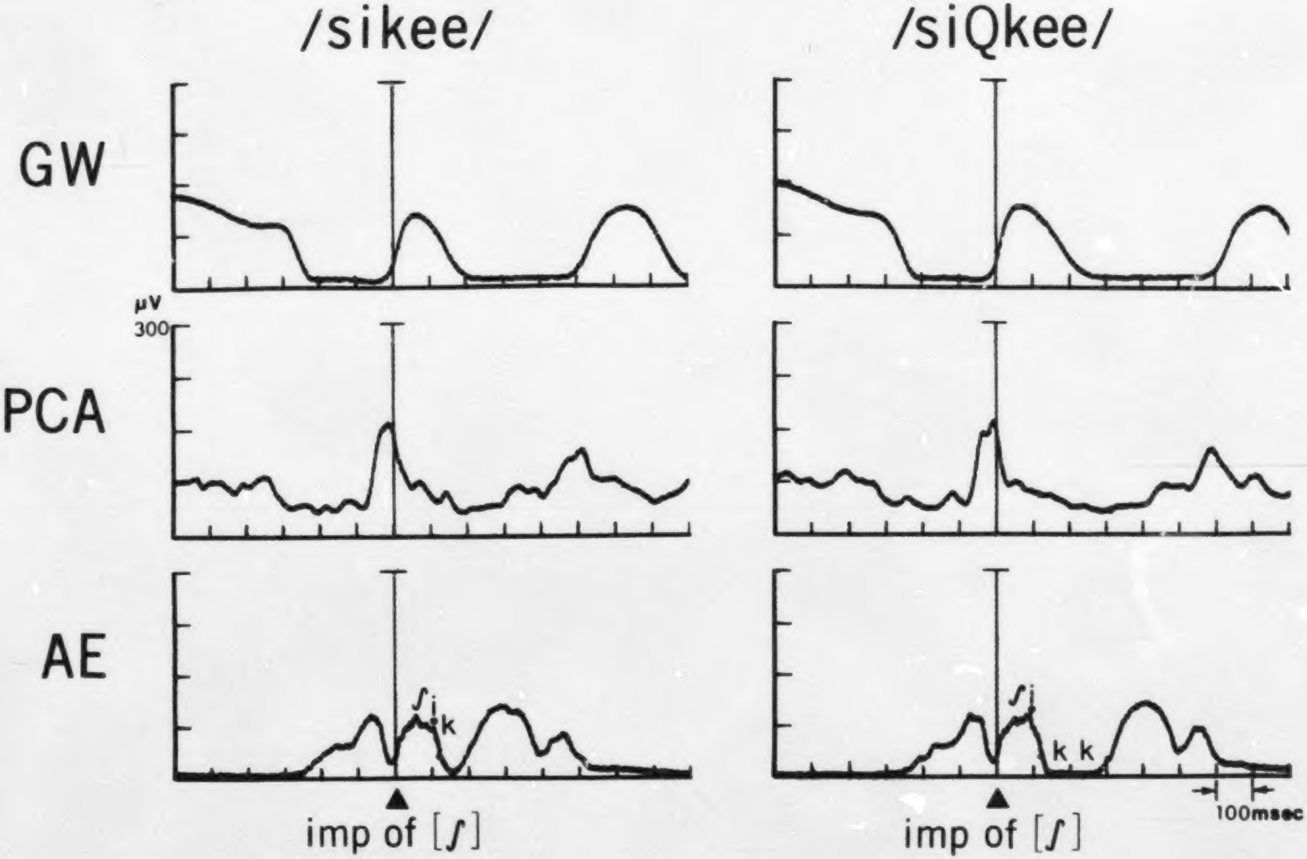


Figure 4. Averaged glottograms, PCA activity patterns and audio envelope curves for the sentences containing the devoiced test words /sikee/ and /siQkee/, respectively.

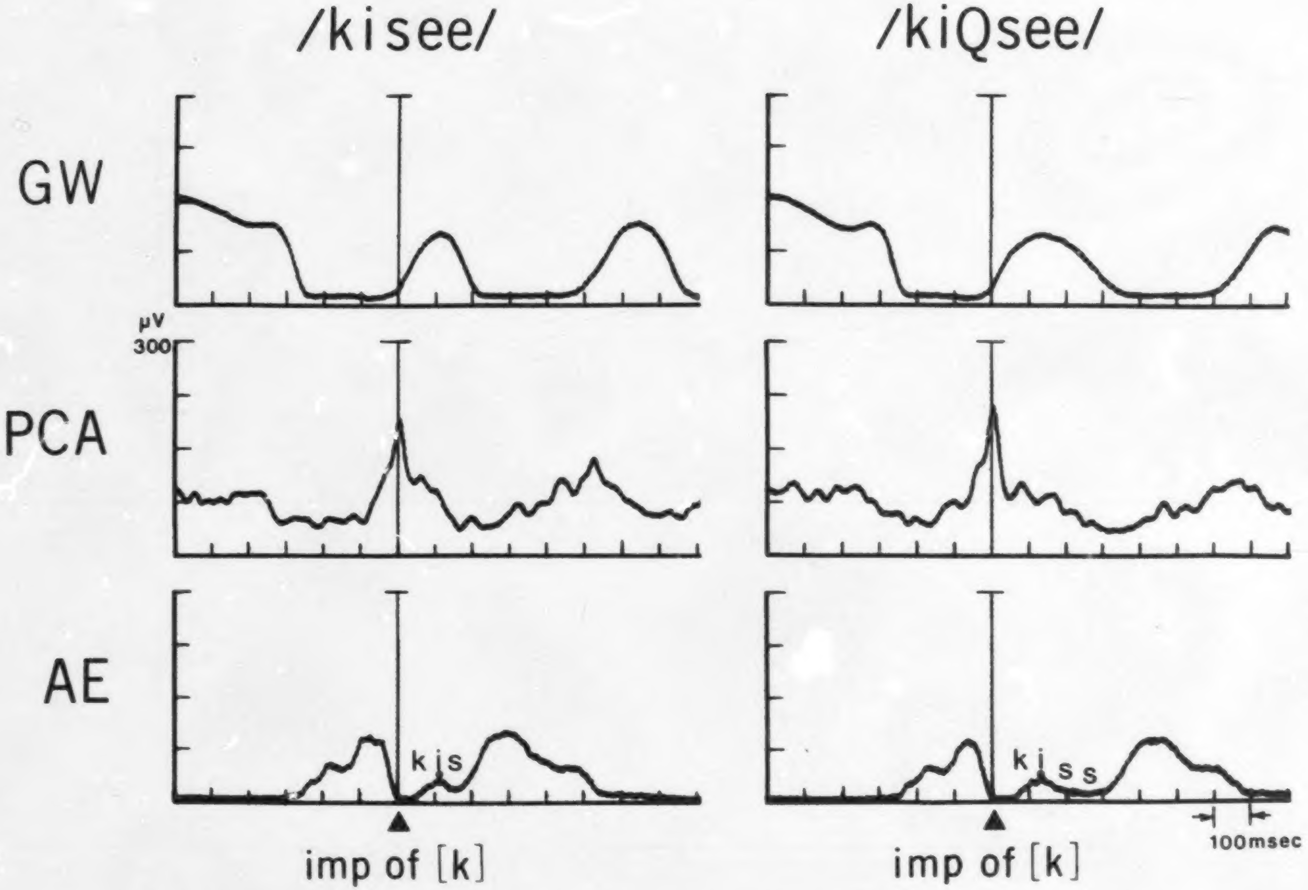


Figure 5. Averaged glottograms, PCA activity patterns and audio envelope curves for the sentences containing the devoiced test words /kisee/ and /kiQsee/. respectively.

obstruents, such as /sik/, /siQk/, /kis/ and /kiQs/. It is evident that all the glottographic curves are characterized by a uni-modal pattern. In addition, and more interestingly, the difference in the slopes during the initial opening phase depends on the phonetic properties of the initial segments: When a voiceless fricative precedes the devoiced vowel, the opening movement is faster than when a voiceless stop precedes the vowel. Furthermore, peak glottal opening during these devoiced sequences coincides approximately with the peak amplitude of the audio envelope signal during the devoiced vowel segments. As for the EMG signals, the noise level is too high for a detailed discussion. Nevertheless, it may be mentioned that the peak PCA activity for these utterance types, as well as for the others mentioned above, occurs around the line-up point, i.e., at the voicing offset of the preceding vowel, regardless of utterance type.

For a detailed comparison of the characteristics of the glottal opening gesture for all the utterance types containing various combinations of voiceless sounds, Figure 6 presents all the glottal movement data superimposed during the pertinent voiceless portions. These averaged curves are again aligned with respect to the voicing offset of the preceding vowel in the frame sentence. The solid lines represent the voiceless sequences beginning with a fricative, while the group of dotted lines corresponds to those beginning with a voiceless stop. The two bottom graphs show separately these two groups, i.e., the sequences beginning with /s/ and /k/, respectively. First of all, with respect to the peak value of the opening gesture, the maximum opening is smaller when the devoiceable vowel is surrounded on both sides by single or geminated stops. In addition, what might be more interesting is that the timing of the peak opening is early and relatively fixed for sequences beginning with fricative /s/, whereas the timing for words beginning with /k/ is comparatively late and more variable than for the /s/ group. Incidentally, it is evident that, except for the word containing /siQs/, these test words may be equally characterized by a single peaked, uni-modal pattern. In other words, only the type /siQs/ is unique, in that the curve has a plateau or two peaks, as stated before. As for the speed of the glottal movement, it seems generally faster for the solid lines, i.e., those for the /s/ group, than for the dotted lines of the /k/ group.

In order to reveal the details of the characteristics of the velocity patterns, Figure 7 shows the velocity patterns for all the utterance types. These plots were made by successive subtractions at 5-msec increments of the glottal width change, using the displacement data in Figure 6. Positive numbers indicate abduction and negative numbers mean adduction. The bottom two graphs are again grouped according to the nature of the initial segments. It is clear that the velocity during the opening phase is faster for sequences beginning with a voiceless fricative than for those beginning with a voiceless stop. Moreover, another interesting finding is related to the timing of peak abduction velocity: The location of peak abduction velocity is almost fixed across both groups, irrespective of the difference in peak amplitudes. Taken together, we may conclude that, although the peak velocity as well as the peak displacement and its timing are clearly different between the /s/ and /k/ groups, the timing of peak abduction velocity is more or less constant in relation to the line-up point, i.e., the voicing offset of the vowel preceding the voiceless sequence.

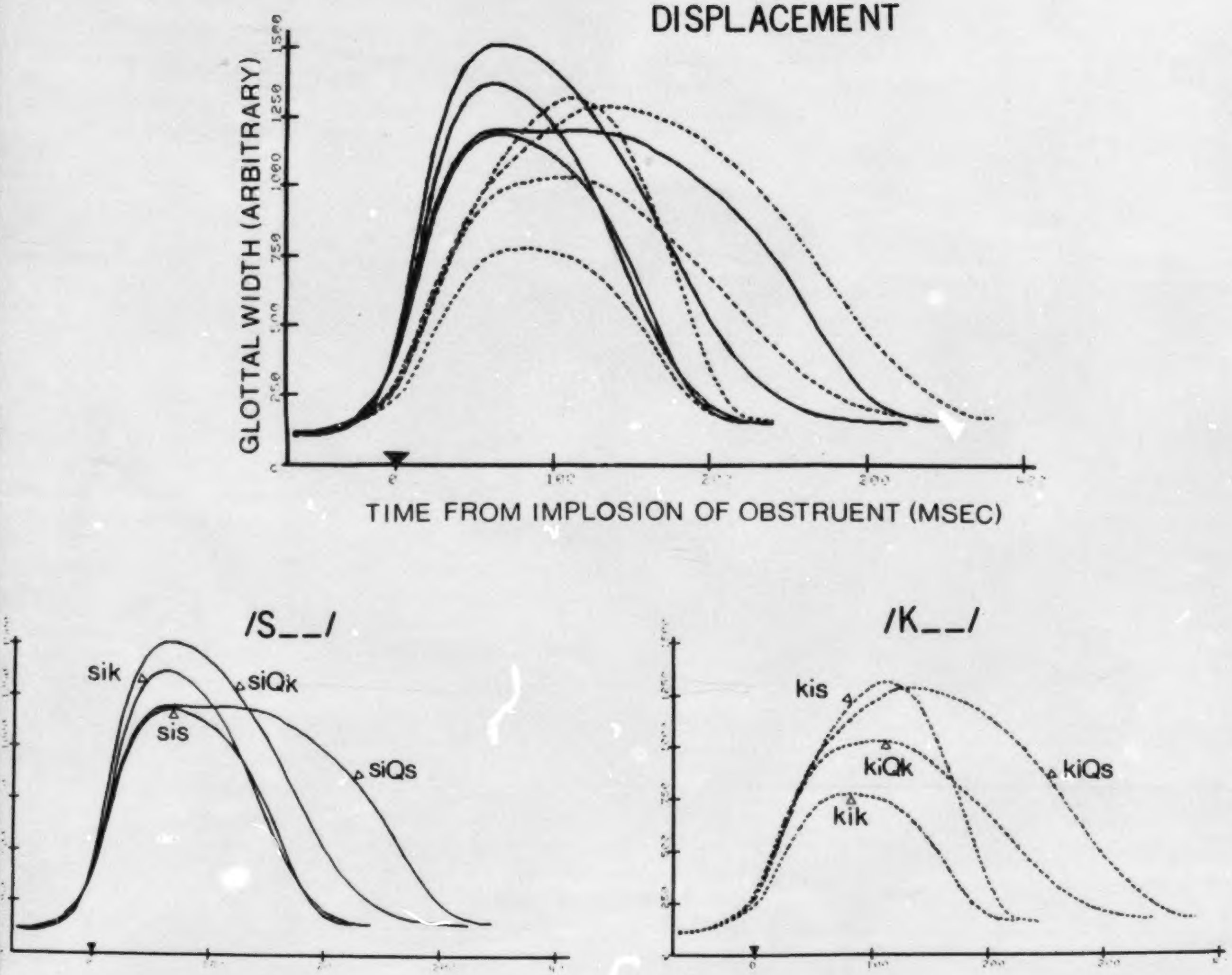


Figure 6. Superimposed curves of the averaged glottograms for all the test voiceless sequences.

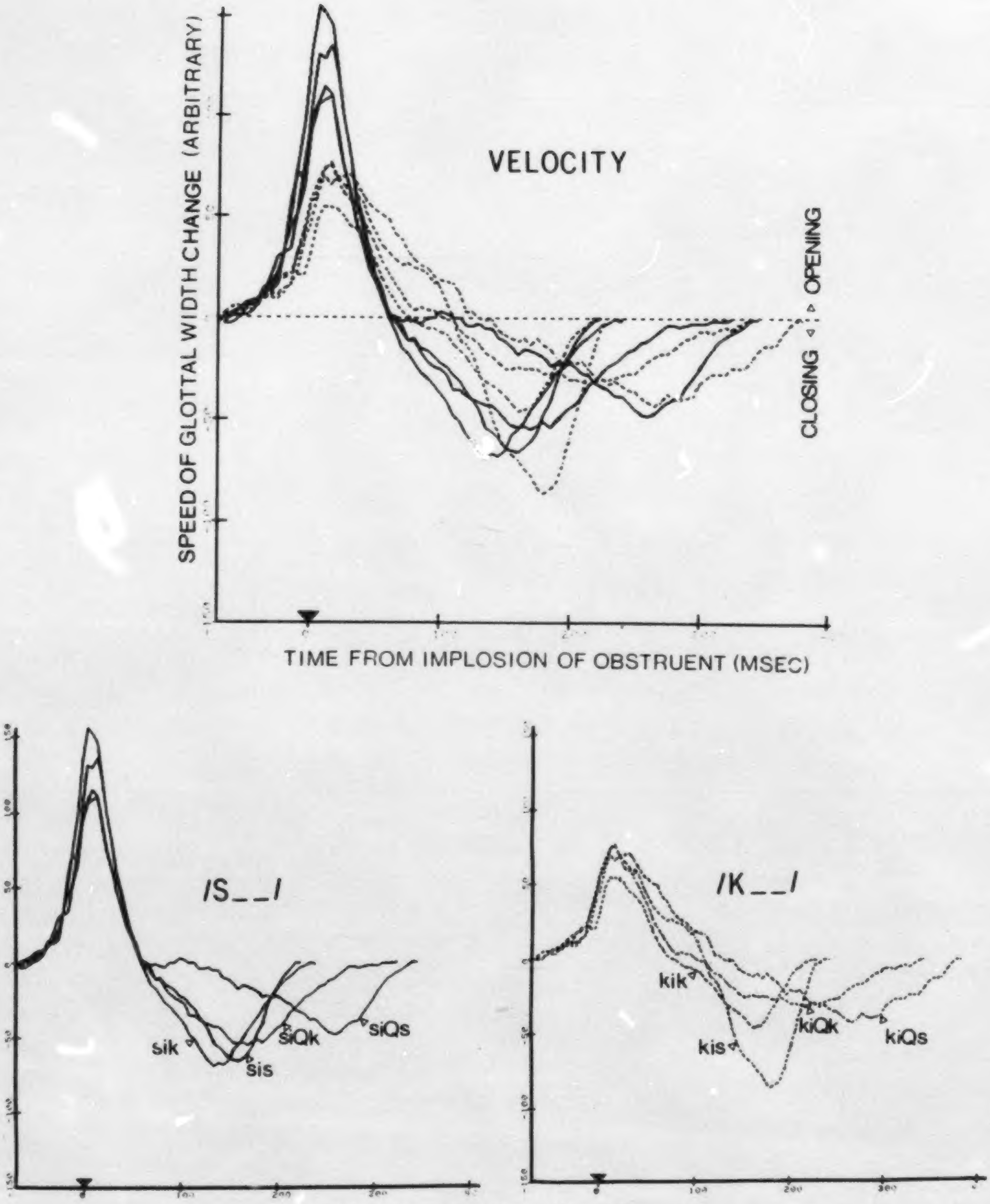


Figure 7. Superimposed curves of the first derivative of the averaged glottograms for all the test voiceless sequences.

DISCUSSION

There are several experiments directed towards understanding the laryngeal adjustments during Japanese voiceless sequence production at both movement and electromyographic levels. For example, Sawashima (1969) showed photoelectric glottograms for single tokens using two native speakers of the Tokyo dialect. According to the data, when the devoiced vowel occurred between two voiceless fricatives, the glottographic patterns tended to be characterized by a slight depression in the middle of the curve for one subject, while the other subject showed a single peaked pattern even in fricative environments. In a later study using fiberoptic filming (Sawashima, 1971), a more comprehensive examination was made, including combinations such as /kis/ and /kik/, which were also used in the present study. He concluded that the fiberoptic data for these voiceless sequences were all characterized by a single peaked curve, although the utterance list did not contain an example of a devoiced vowel surrounded on both sides by voiceless single and/or geminated fricatives. Recently, Sawashima and his colleagues have reported on simultaneous fiberoptic and electromyographic recordings for single tokens, showing a two peaked pattern during /siQs/ sequence production for two subjects (Sawashima, Hirose, & Yoshioka, 1978).

The present results, although limited to a single subject and presented as ensemble-averages, appear to be generally in good agreement with these previous works: When the devoiced vowel occurs between a voiceless fricative and a geminated one, such as /siQs/, the glottal opening gesture may be characterized by a bimodal or, at least, a plateau-type pattern. In contrast, all the other glottal opening patterns during voiceless sequence production are characterized by a rather simple, single peaked pattern. Of course, it should be taken into consideration that these findings might reflect speaker-specific and/or token-specific aspects (e.g., Sawashima, 1969). Nevertheless, it is always found, and also in the other studies of Japanese, that a voiceless fricative environment, and typically the one containing a geminate, seems to require two separate opening gestures.

In addition, the current data also reveal the detailed characteristics of the averaged photo-electric glottograms, demonstrating the dependence of the abduction gesture on the phonetic nature of the segments: When the voiceless sequence contains a voiceless fricative /s/, the peak value of the glottal opening is larger than that for the one without a fricative. Moreover, the timing of the first peak opening varies according to the property of the initial segments: Early and relatively fixed for the fricative initial group, and late and more variable for the stop initial group. This finding is also consistent with our recent studies using American English (Yoshioka et al., 1979), Icelandic (Löfqvist & Yoshioka, in press), and Swedish (Löfqvist & Yoshioka, 1980), although the phonologies of these languages differ, among other things, in the significance of stop aspiration. Therefore, we are inclined to conclude that at least the difference in the peak value between a voiceless fricative and a voiceless stop is universal.

Furthermore, the plots of the velocity curves add another new dimension: Despite the clear difference of the peak value of the velocity between stop and fricative initial groups, the timing of the peak of the abduction velocity is almost fixed across the two groups. It should be mentioned here that the line-up point was determined as the voicing offset of the preceding vowel regardless of the nature of the initial voiceless segment. In considering the fact that the glottis is usually slightly open at this moment, in particular when the initial segment is a voiceless fricative, peak velocity for the frictive initial group might occur a little later than that for the stop group, if the beginning of the opening movement, defined as the inflection point in the movement curve, was chosen as the line-up point.

These results, in conjunction with other studies of ours using different languages mentioned above, may be interpreted in several ways. From a phonetic viewpoint, the faster and larger opening for a voiceless fricative may be related to the necessary supply of air during the voiceless fricative segment to produce adequate turbulent noise by a quick reduction of laryngeal resistance. On the other hand, in order to stop glottal vibrations at the implosion of a voiceless stop and assist in the buildup of oral pressure, a slight opening gesture may be sufficient in combination with the the closing gesture of oral articulators. As for the fixed timing of the peak abduction velocity across different phonetic sequences, the interpretation seems open. From a physiological aspect, however, it is possible that this fixed timing reflects a basic nature of the voluntary movement control of the glottis particularly in relation to oral gestures: It could be that the timing of velocity is physiologically constrained, while the magnitude of velocity and displacement are adjusted within such a temporal framework to meet various phonetic requirements.

REFERENCES

Basmajian, J., & Stecko, G. A new indwelling electrode for electromyography. Journal of Applied Physiology, 1962, 17, 849.

Fujimura, O. Hatsuon no katei. Japan Journal of Logopedics and Phoniatrics, 1971, 12, 11-21.

Han, M. S. Unvoicing of vowels in Japanese. Study of Sound, 1962, 10, 81-100.

Hattori, S. Onseigaku. Tokyo: Iwanamishoten, 1951.

Hirano, M., & Ohala, J. Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. Journal of Speech and Hearing Research, 1969, 12, 362-373.

Hirose, H. Posterior cricoarytenoid as a speech muscle. Annals of Otology, Rhinology and Laryngology, 1976, 88, 334-343.

Hirose, H., Gay, T., Strome, M., & Sawashima, M. Electrode insertion technique for laryngeal electromyography. Journal of the Acoustical Society of America, 1971, 50, 1449-1450.

Kewley-Port, D. EMG signal processing for speech research. Haskins Laboratories Status Report on Speech Research, 1977, SR-50, 123-146.

Löfqvist, A., & Yoshioka, H. Laryngeal activity in Swedish obstruent clusters. Journal of the Acoustical Society of America, 1980, 68, 792-801.

Löfqvist, A., & Yoshioka, H. Laryngeal activity in Icelandic obstruent production. To appear in Nordic Journal of Linguistics, in press.

Sawashima, M. Devoiced syllables in Japanese: A preliminary study by photoelectric glottography. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1969, *3*, 35-41.

Sawashima, M. Devoicing of vowels. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1971, *5*, 7-13.

Sawashima, M. Hatsuonji no koto chosetsu. In S. Hiki (Ed.), *Onseikagaku*. Tokyo: Tokyo University Press, 1973.

Sawashima, M. Current instrumentation and technique for observing speech organs. *Technocrat*, 1976, *9-4*, 19-26.

Sawashima, M., & Hirose, H. A new laryngoscopic technique by use of fiberoptics. *Journal of the Acoustical Society of America*, 1968, *43*, 168-169.

Sawashima, M., Hirose, H., & Yoshioka, H. Abductor (PCA) and adductor (INT) muscles of the larynx in voiceless sound production. *Annual Bulletin* (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 1978, *12*, 53-60.

Yoshioka, H. Laryngeal adjustments during Japanese fricative and devoiced vowel production. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-58*, 147-160.

Yoshioka, H., Löfqvist, A., & Hirose, H. Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Haskins Laboratories Status Report on Speech Research*, 1979, *SR-59/60*, 127-151.

ARTICULATORY CONTROL IN A DEAF SPEAKER*

Nancy S. McGarr+ and Katherine S. Harris++

INTRODUCTION

While many children who are born severely or profoundly deaf, or become deaf in infancy achieve intelligible speech, the vast majority do not. Speech intelligibility is fairly well correlated with residual hearing (Boothroyd, 1970; Smith, 1972) at least until 90dB, and overall intelligibility is well correlated with the percent of segmental errors, and to a lesser extent with suprasegmental deviancy (Levitt, Smith, & Stromberg, 1974). While many educators of the deaf would claim that the characteristic unintelligibility of deaf speakers is a consequence of faulty teaching practices (Haycock, 1933; Ling, 1976), independent investigations have been remarkably consistent in showing similar patterns of segmental and suprasegmental errors in the speech of deaf talkers trained in a wide variety of programs (Hudgins & Numbers, 1942; Smith, 1972; Levitt, Stark, McGarr, Carp, Stromberg, Gaffney, Barry, Velez, Osberger, Leiter, & Freeman, Note 1; Johnson, 1975). Furthermore, experienced teachers of the deaf can discriminate between deaf and non-deaf speakers from disyllables produced by both groups (Calvert, 1961), and experienced listeners of the deaf are better than naive listeners in decoding deaf utterances (McGarr, 1978). If we accept the point of view that there is a generic "deaf speech"[1] pattern, not dependent at least on the fine-grained details of the training procedure, we may ask what are its characteristics? Why do the deaf sound as they do? Why are they unintelligible?

One hypothesis, primarily concerned with consonant articulation, is that deaf speakers place their articulators fairly accurately--especially for those places of articulation that are highly visible--but fail to coordinate the movements of several articulators normally (Huntington, Harris, & Sholes, 1968; Levitt et al., 1974). Thus, we may suggest that the errors in deaf speech are the consequences of incorrect motor planning in time.

---

*To appear in Hochberg, I., Levitt, H., and Osberger, M. J. (Eds.), _Speech of the hearing impaired: Research, training and personnel preparation_. Washington, D.C.: A. G. Bell Association, in press.

+Also Molloy Catholic College for Women, Rockville Center, N.Y.

++Also Graduate School and University Center, The City University of New York.

Acknowledgment: The acoustic measurements were described in a paper presented at the meeting of the Acoustical Society of America, Atlanta, Georgia, April 1980. We are grateful to our colleagues Fredericka Bell-Berti and Carole E. Gelfer for their helpful comments and assistance. The work described in this paper was supported by Grants NS-13617, NS-13870, and RR-05596 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

A second hypothesis, primarily concerned with vowel articulation, is that deaf speakers move their articulators through a relatively restricted range, thereby "neutralizing" vowels (Angelocci, Kopp, & Holbrook, 1964; Monsen, 1974). However, this hypothesis fails to account for the great variability in the speech production of deaf talkers, a point we will discuss in further detail later.

A third hypothesis is that the inability of deaf speakers to control the suprasegmental characteristics of their speech makes both segmental and suprasegmental characteristics more difficult for listeners to decode (Harris & McGarr, 1980). Suprasegmental aspects of speech may be so abnormal as to mislead the listener. Deaf speakers may not preserve phonological contrasts or may produce them in a way that makes information about the intended contrast unavailable to the listener, and perhaps block information about other contrasts. That fundamental frequency (McGarr & Osberger, 1978) and overall duration levels (e.g., Osberger & Levitt, 1979) are often deviant in deaf speakers is well known. These deviations alone might interfere with a listener's ability to decode a speech signal, even if other suprasegmental contrasts were preserved in either a normal or an abnormal way.

On an entirely different level, poor control of the speech source function may simply provide inadequate support for the acoustic realization of upper articulator movement. Deaf speakers characteristically take in less air in speech respiration (Forner & Hixon, 1977; Whitehead, in press) and may, in addition, convert air into acoustic energy inefficiently due to poor control of the larynx.

This paper presents a preliminary attempt to assess these hypotheses by examining a number of productions of some simple utterances by a single deaf talker using listeners to judge production accuracy utterance-by-utterance. While it is obvious that more subjects must be studied in order to reach firm conclusions, we believe that the general technique of examining interarticulator programming in depth with combined perceptual, acoustic, and physiological techniques is a promising avenue for investigation.

METHODS AND PROCEDURES

The prelingually deaf speaker in this study is a woman in her mid-forties who graduated from an oral school for the deaf, and has received remedial speech classes as an adult. Her pure tone average is 105dB ISO. Informal ratings of spontaneous speech samples suggest that her productions would be characterized as fairly typical of her group. For purposes of comparison, productions of a hearing speaker who has frequently served as an experimental subject were also examined.

Each subject produced approximately 20 repetitions of each of six utterance types. These utterances were nonsense words of the type /əpipɑp/, /əpipip/, and /əpɑpip/ with stress on either the /i/ or the /ɑ/. For this paper, data will be presented primarily for the first and third utterance types. Paint-on surface electrodes were used to record from the orbicularis oris muscle (Allen, Lubker, & Harrison, 1972); conventional hooked-wire electrodes were inserted into the genioglossus muscle. The electrode preparation and insertion techniques for the genioglossus muscle electrodes have been reported in detail elsewhere (Hirose, 1971). Conventional acoustic recordings were made at the same time as the electromyography.

The acoustic and electromyographic (EMG) data obtained from the two speakers were analyzed in several ways. First, for the deaf speaker, the acoustic recordings of six utterance types were randomized and presented to listeners inexperienced in hearing deaf speech. The listeners were required to select one of the six utterance types presented on an answer sheet, for each item they heard. Confusion matrices were obtained. The hearing subject's productions were not checked perceptually, but informal listening suggested that perceptual errors would not be made by listeners to her speech. Second, acoustic measurements were made on an interactive computer system at the Haskins Laboratories and with conventional sound spectrography. Third, the EMG signals were rectified, integrated, and then further analyzed, as we will describe below.

RESULTS

Listener Judgments

First, examining the results of the listening test, we found that the deaf speaker was judged as being fairly intelligible (at least as measured by a closed response listening task). Table 1 shows the confusion matrix obtained from the listeners' scores. An item was considered to be correct if 9 out of 10 listeners identified it as the originally intended utterance. The average percent correct for all utterance types was 75%. Overall, there were more errors of stress than of the segment type (i.e., a vowel identity error). In fact, only for the utterance /ə pɑ'pip/ was there a significant number of vowel errors. In this case, the listeners perceived the utterance as /ə pipip/ 32% of the time.

Table 1

Confusion Matrix of Listeners' Judgments for the Deaf Speakers

	1	2	3	4	5	6
1. ə'pi pap	88	08				
2. əpi'pap	25	75				
3. ə'pi pip			83	17		
4. əpi'pip			07	91		
5. ə'pa pip					67	29
6. əpa'pip				32	16	51

Using these listener judgments, all tokens (repetitions) of an item were divided into two categories: "perceived correct" utterances and "stress error" utterances. Only for the intended utterance /ə pɑ́pip/ was there an additional category (that of a vowel error).

Acoustic Measurements

The acoustic cues used to convey contrastive stress in normal speech production have been extensively studied (Fry, 1958, 1964; Harris, 1978). In general, speakers convey changes in contrastive stress to listeners by differences in acoustic cues such as vowel duration, fundamental frequency, amplitude, and formant frequency. For the deaf speaker, two questions are of interest. First, what acoustic cues does a deaf speaker use to convey contrastive stress to the listener and how do these cues compare to those used by the normal speaker? Second, can productions perceived as being incorrect in the speech of the deaf be explained as differing systematically from those utterances perceived as being correct?

If stress may be conveyed at least in part by differences in vowel duration, we might expect that for "perceived correct" utterances in the speech of the deaf, the stressed vowel would be longer than the unstressed vowel. Conversely, "stress error" utterances may be due, in part, to an inappropriate vowel duration ratio.

The measurements of vowel duration show that the deaf speaker was like the hearing speaker in some ways, but not in others. Figure 1 shows the measurements of vowel duration for the hearing speaker (FBB) and the deaf speaker's "perceived correct" utterances (MH↑) and "stress error" utterances (MH↓). Dark bars represent stressed vowels; open bars represent unstressed vowels. As expected, overall duration of the vowels produced by the deaf speaker was considerably longer than that of the hearing speaker.

For the hearing speaker, there is always a shift towards longer relative duration for a vowel when it is stressed than when it is not, although this pattern is apparently complicated by differences in intrinsic vowel duration in that productions of /ɑ/ are in general longer than productions of /i/ in the same phonetic environment. An acoustic analysis of a second hearing speaker shows less effect of intrinsic vowel duration. However, the deaf speaker did not show consistent differences in intrinsic vowel duration between /i/ and /ɑ/ within the same phonetic context.

On average, the deaf speaker appears to be conveying contrastive stress by varying vowel duration in the sense that intended stressed vowels were always longer than unstressed vowels in the same utterance, and across utterances. For example, in the utterance ′i-ɑ, when perceived as intended (↑), the average duration of /i/ was 334 msec; in the contrastive pair /i-′ɑ/, when /i/ was not stressed, its duration was 267 msec. The same pattern--stressed vowels longer than unstressed--holds for all vowels perceived as correct. However, we find nearly the same pattern for "stress error" utterances. That is, when an unstressed /i/ was perceived in the first contrast /′i-ɑ/, the duration of the /i/ was 380 msec, and when a stressed /i/ was perceived in the contrastive pair /i-′ɑ/, the /i/ was 285 msec. Thus, the same pattern of vowel durations was found in both "perceived correct" and "stress error" utterances.

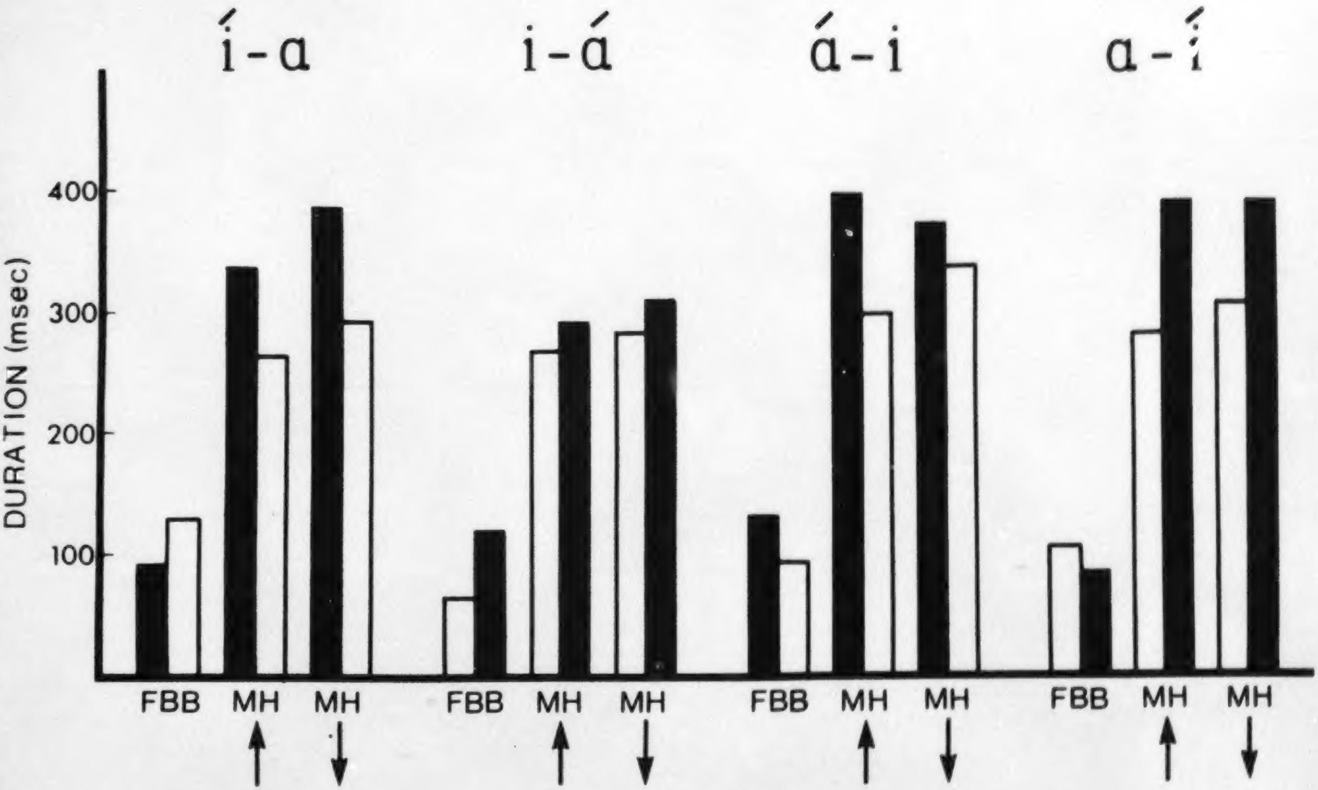


Figure 1. Mean duration of vowels for the hearing speaker and the deaf speaker.

In Figure 2, the data show the mean vowel durations and their standard deviations. The durations of the hearing speaker's utterances show very little variability, as reflected in the small standard deviations. In contrast, the deaf speaker was exceedingly variable. Standard deviations were fairly large for the deaf speaker and vowel durations for correct and incorrect utterances often fell within the same range.

The data in Figures 1 and 2 suggest that the deaf speaker is not conveying stress contrasts primarily by differences in vowel duration and also that perceived stress errors are not due simply to a consistently used incorrect pattern of duration. Instead, it would seem that the deaf speaker learned the stress rules of relative vowel duration but is unable to use them to produce an acoustically constant output.

Figure 3 shows measurements of fundamental frequency (F_0) obtained from extracting individual pitch periods from the middle portion of each vowel and calculating the frequency from the period. In making these measurements, we noted frequent abnormalities of the waveform. For the hearing speaker, F_0 is higher for stressed than for unstressed vowels, as expected. For the deaf speaker, F_0 is higher for the intended stressed vowel in three of the four utterance types, but for /ˈɑ-i/, F_0 is slightly lower for the intended stressed vowel in both "perceived correct" and "stress error" utterances. Again, as with duration, patterns are the same for "perceived correct" and "stress error" utterances.

In Figure 4, the data show mean F_0 and its standard deviation. For the hearing speaker, the standard deviations are small, again reflecting little variability. Obviously, the standard deviations for the deaf speaker are large, indicating that the utterances were quite variable. Again, these data suggest that perceived errors are not due simply to a consistently used incorrect pattern of F_0.

Figure 5 shows measurements of the amplitudes of the vowels relative to a standard, the first production of an unstressed /ɑ/ in the utterance /əˈpipɑp/. For the hearing speaker, not surprisingly, stressed /ɑ/ had greater amplitude than stressed /i/ and the amplitude of a given vowel increased with stress. For the deaf speaker, the stressed vowel always had a higher amplitude than the unstressed vowel. But again, it is clear that this deaf speaker is not conveying contrastive stress to the listener by differences in relative amplitude since "correct" and "incorrect" productions show the same pattern.

Another way in which stress change may be conveyed acoustically is by differences in vowel color. Fry (1964) has shown that listeners are more likely to perceive a syllable as unstressed if the formant values are less extreme, or more like the neutral schwa. Physiological explanations for the effect have been proposed by Lindblom (1963) and by Harris (1978). Without going into the details, it should be noted that the Harris study included measurements of productions of the same disyllables by the same speaker, FBB. We therefore measured the values for the deaf speaker, as presented in Table 2. The results show neither a consistent pattern overall, nor a systematic difference between "correct" and "incorrect" utterances. However, it should be noted that measurements were extremely difficult to make either because of

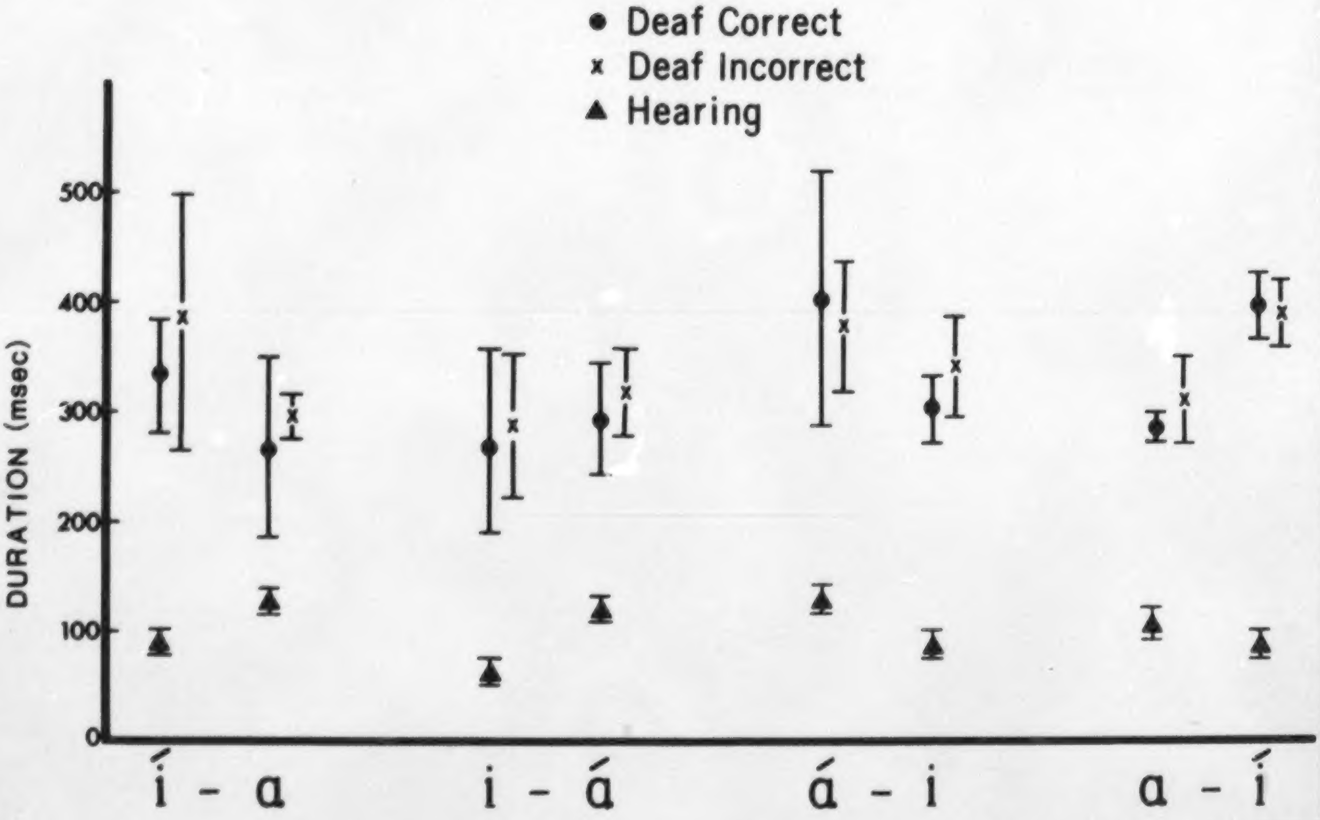


Figure 2. Mean and standard deviations of vowel duration for the hearing and deaf speaker.

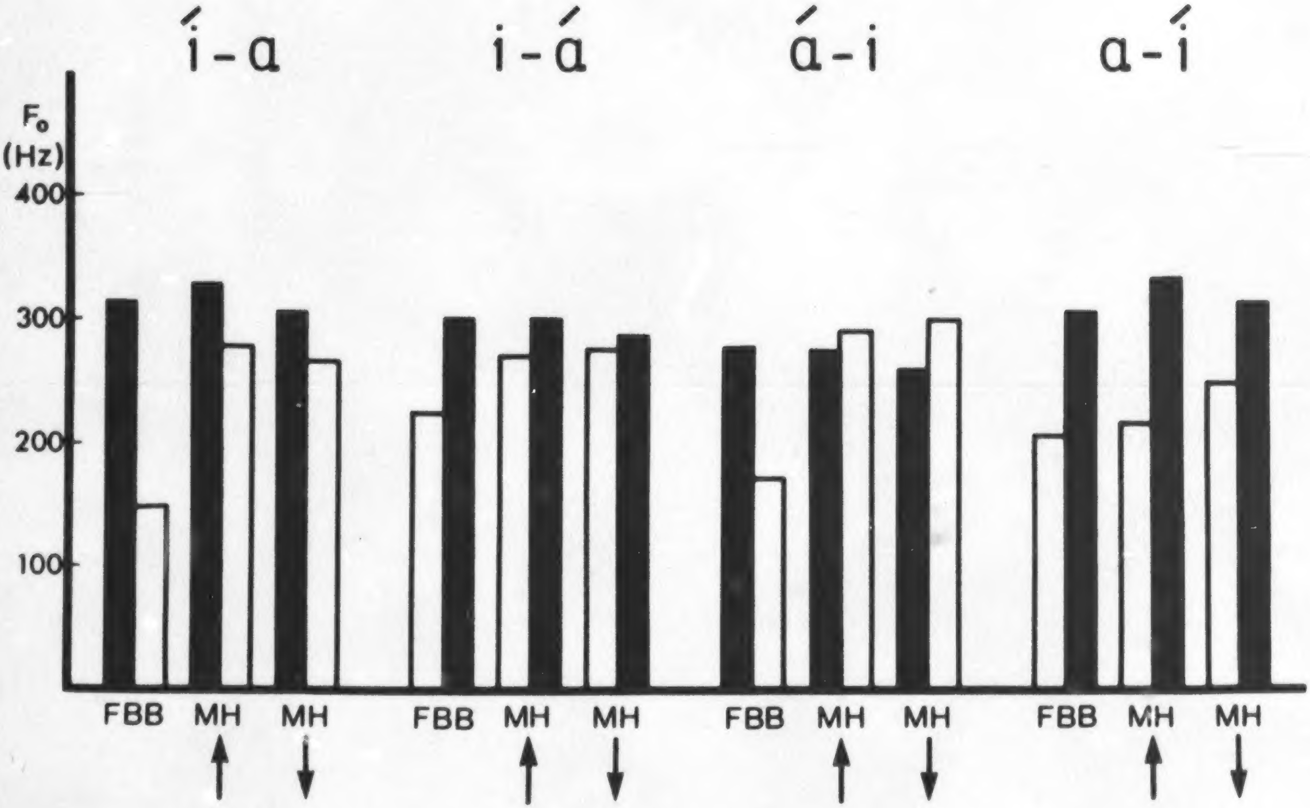


Figure 3. Mean fundamental frequency (F_0) for the hearing and deaf speaker.

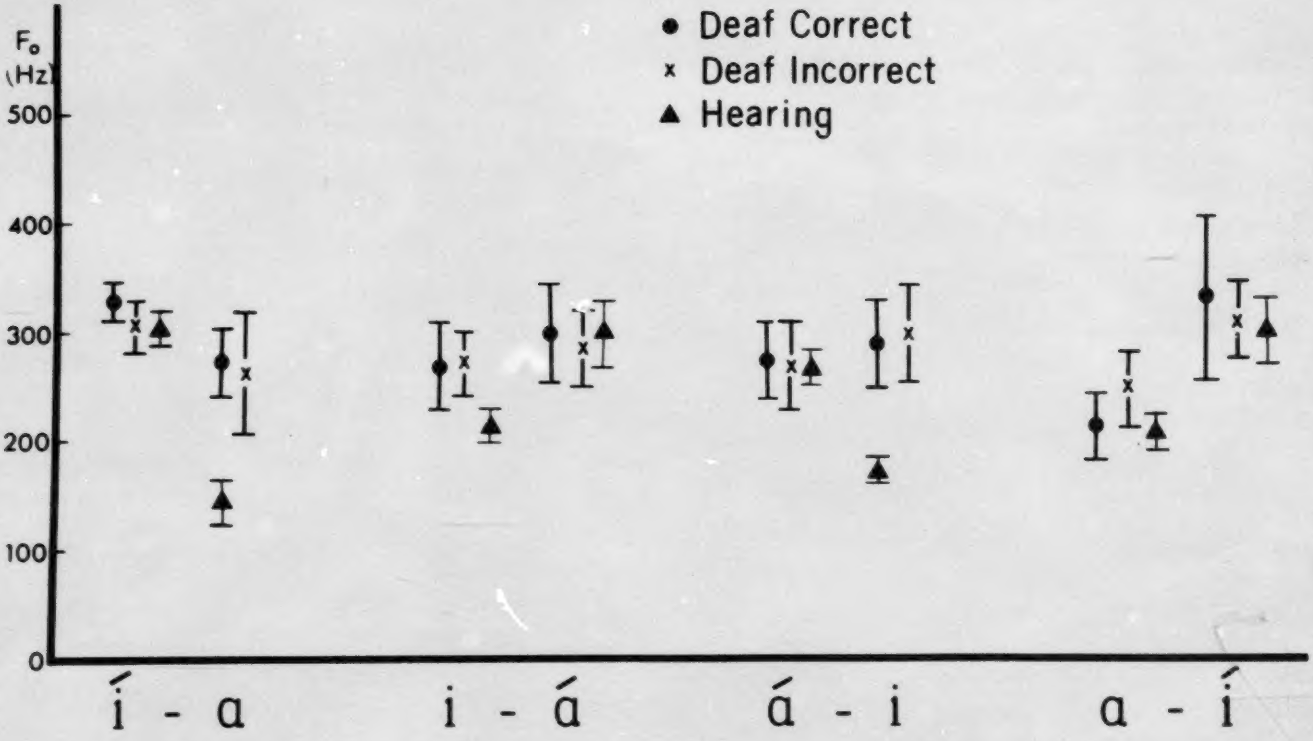


Figure 4. Mean and standard deviations of F_0 for hearing and deaf speakers.

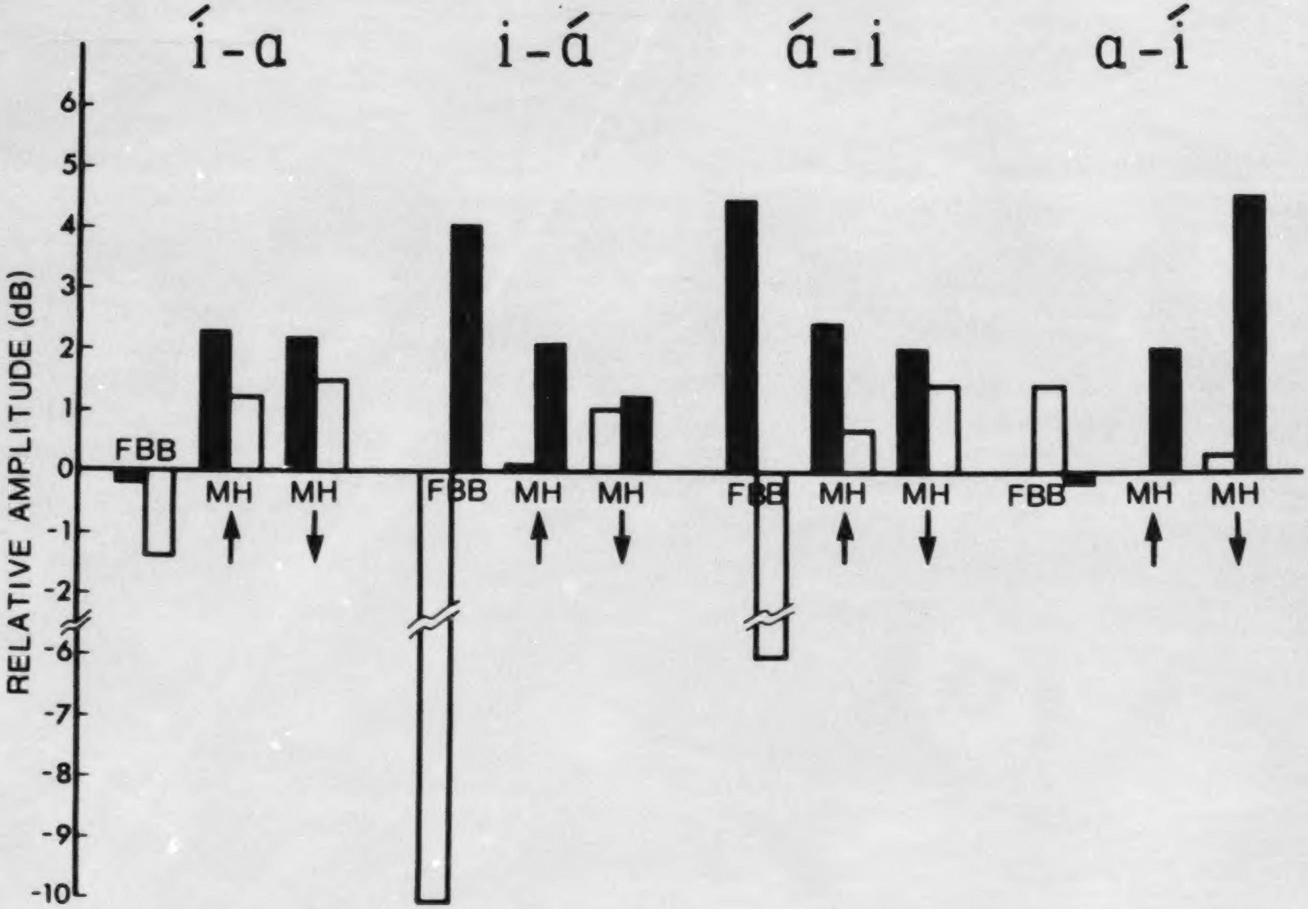


Figure 5. Mean relative amplitude (dB) of the vowels for the hearing and deaf speakers. The standard was the first production of an unstressed /ɑ/ in the utterance /ə'pipɑp/.

Table 2

Mean Values for F_2 and F_3 for the Deaf Speaker's Utterances Perceived Correct or Perceived Incorrect

	F_2	F_3	F_2	F_3
1. ə́ pi pap	i		a	
Correct	2170	2990	1546	2369
Incorrect	2060	2940	1500	2330
2. əpiˈpap				
Correct	2162	2950	1625	2475
Incorrect	2170	2880	1670	2370
3. ə́ pi pip	i		i	
Correct	2188	3055	2066	2766
Incorrect	2190	3060	2110	2950
4. əpiˈpip				
Correct	2246	2980	2280	3100
Incorrect	2200	2900	2166	3100
5. ə́pa pip	a		i	
Correct	1620	2600	2040	2880
Incorrect	1550	2592	2150	2875
6. əpaˈpip				
Correct	1733	2600	2100	2966
Incorrect	1650	2320	2100	2970

the mismatch between spectrograph filter and fundamental frequency (cf. Huggins, 1980), or because of source function abnormalities.

This deaf speaker appears, at least on average, to have learned some rules for conveying stress increase: vowel duration longer, F_0 higher, and amplitude higher. Furthermore, it is not likely that these were specifically included in this deaf speaker's training program since theoretical discussions of suprasegmental production at this level are relatively recent in the literature on training deaf speakers. More likely, this speaker has extracted this information from her low frequency residual hearing and then generalized it to abstract rules. However, the variability in her production suggests an inability to coordinate the production mechanism so as to achieve these stress contrasts in a consistent acoustic manner. Furthermore, although she communicates the information that should allow listeners to judge stress, they evidently cannot use it.

EMG Results

The electromyographic (EMG) results were examined to see if they revealed any systematic differences between normal and deaf interarticulator programming, or between correctly and incorrectly perceived utterances. In these utterances, orbicularis oris (OO) activity is associated with pursing and closing the lips as for the /p/. For the vowel /i/, the genioglossus (GG) bunches the tongue and brings it forward in the mouth (Raphael & Bell-Berti, 1975; Raphael, Bell-Berti, Collier, & Baer, 1979).

Figure 6 shows data for the hearing speaker producing the utterance type /ə́ papip/. At the top of each column (genioglossus at the left, orbicularis oris at the right) is the ensemble average of the EMG waveforms. This was obtained by rectifying and integrating the EMG potentials for each repetition and aligning them with respect to an acoustic event. The signals were digitized and the ensemble average calculated by averaging each sample for each repetition of an utterance type (Kewley-Port, 1973). A sample of four of the 20 repetitions is seen in the columns below the average. For this utterance type, the line-up point for averaging the EMG and acoustic events, indicated by the vertical line at 0 msec, is the release burst of the second /p/.

The data for orbicularis oris show three well-defined peaks of activity corresponding to the lip gestures for the three /p/ closures in /ə́pɑpip/. The line-up point falls between peaks 2 and 3. The duration of the interval between peaks 1 and 2 is greater than that between peaks 2 and 3, reflecting the longer duration of the /ɑ/. One notable feature of these data is the striking similarity of the EMG patterns for all tokens. For the genioglossus, there is a peak of activity for the /i/ and no activity for the /ɑ/ as expected, since the genioglossus is active in raising and bunching the tongue. Indeed, peak genioglossus activity (for the vowel) occurs approximately at the time of the acoustic line-up event--the /p/ burst-release. This is not suprising since EMG activity precedes the articulatory event to which it is related by about 50-100 msec.

Figure 7 shows data for the utterance /ə pɑ́ pip/ again for the hearing speaker. The interval between the second and third peaks of orbicularis oris

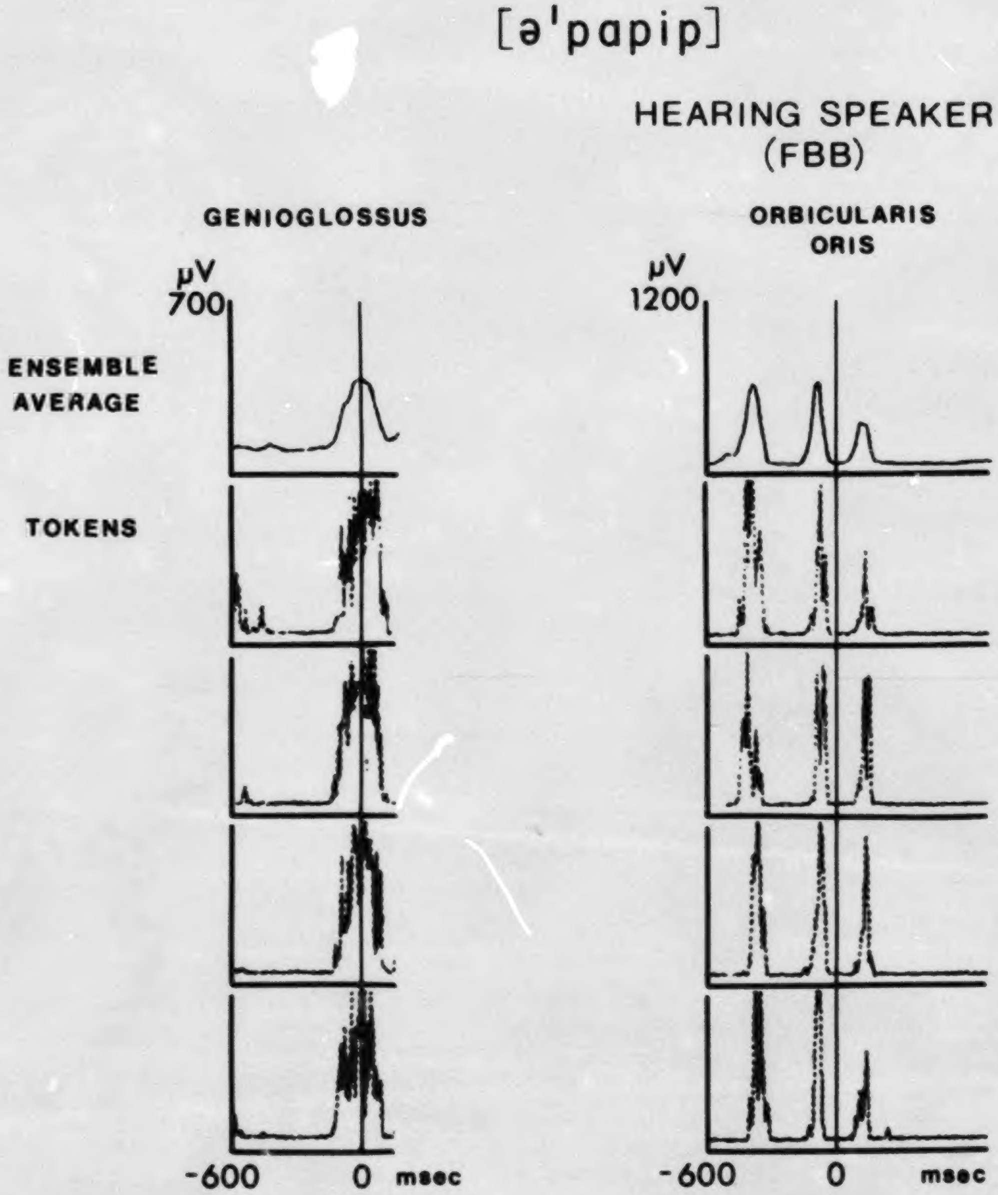


Figure 6. /ə'pɑpip/ as produced by a hearing speaker. Data plots at the top show the EMG averaged for about 20 tokens for the genioglossus and orbicularis oris muscles. Four individual tokens are shown below. The vertical line indicates the acoustic release of the /p/ closure.

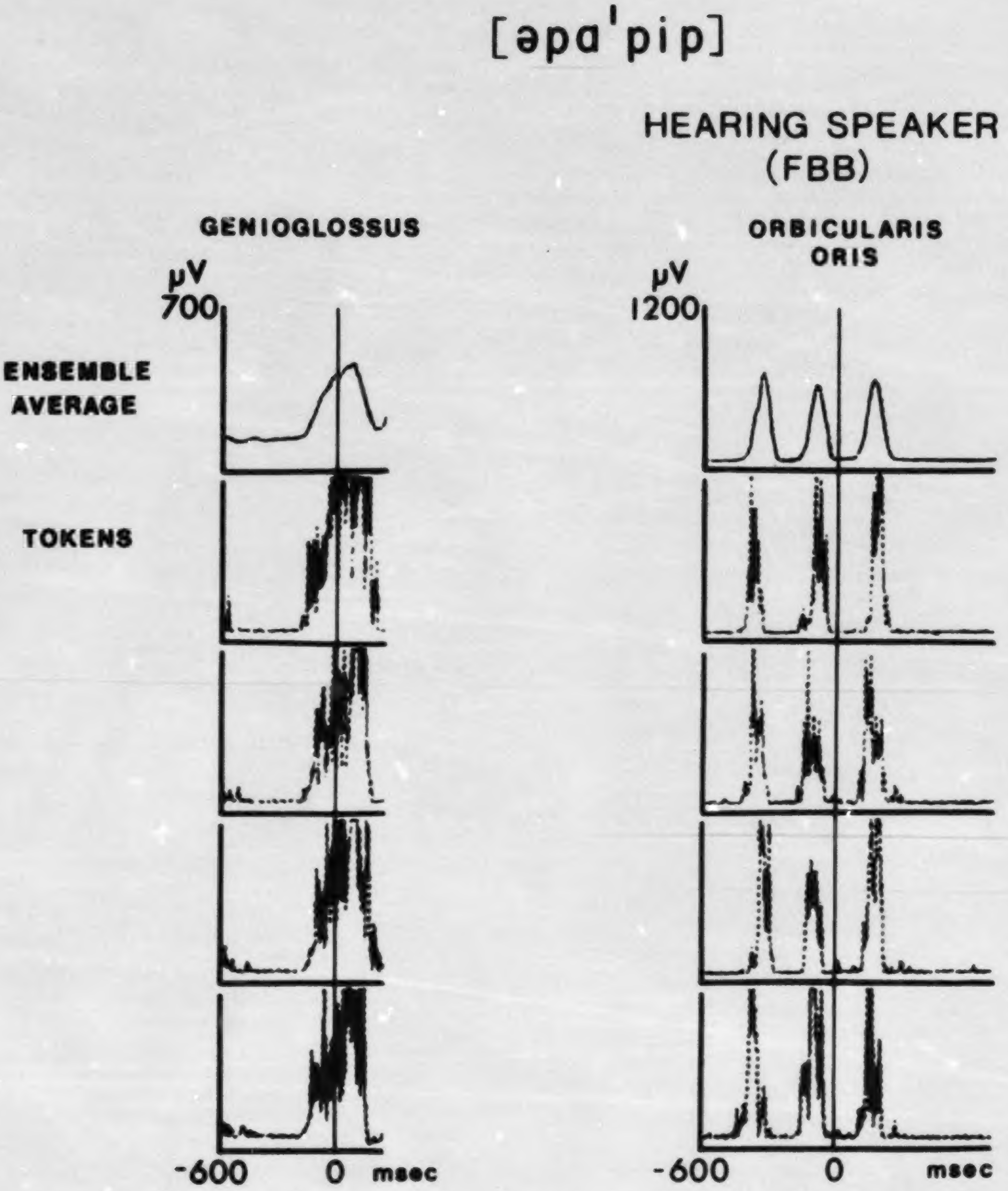


Figure 7. /əpápip/ as produced by a hearing speaker. Data presented as in Figure 6.

activity is greater than that between the first and second peaks since the vowel in the final syllable is longer. Also, the duration of genioglossus activity is longer in this utterance type, since /i/ is stressed. Note, however, that peak activity for the genioglossus still occurs at the release of the second /p/, between peaks 2 and 3. Once again, the pattern of activity for all these tokens looks remarkably similar.

Figure 8 shows parallel data for several of the deaf subject's productions of /ə'p ɑpip/. Each of these tokens was a "perceived correct" utterance. Examining the EMG activity for orbicularis oris we see that, as for the hearing subject, there are three well-defined peaks of activity and the interval between the second and third peaks is greater than that between the first and second peaks. However, the duration of each peak is prolonged. The /p/ release falls between the second and third peaks as for the hearing speaker.

Turning to the genioglossus EMG, peak activity is less well defined and occurs later than for the hearing speaker; it follows /p/ release. Further, there is considerable variability from token to token in the duration of genioglossus activity. In some instances, this activity starts fairly early (token 3) and at other times, later (token 4).

Figure 9 shows the data for the deaf speaker's production of /ə pɑ'pip/. Here again, the overall duration of EMG activity is prolonged for both muscles, but the pattern more closely resembles that of the hearing speaker for orbicularis oris than for genioglossus. The variability and "lateness" of the genioglossus are again observed. These data show that the deaf speaker was somewhat like the hearing speaker with respect to "the visible aspects of articulation," but quite variable with respect to the timing of lingual control. This variabiltiy appears to be particularly manifested in what we would describe as abnormal interarticulator coordination. To illustrate this notion further, the data for selected tokens of orbicularis oris and genioglossus were plotted.

For purposes of comparison, Figure 10 shows the averaged EMG activity for these muscles for the hearing speaker. Onset of the genioglossus activity is closely coordinated with the second peak of orbicularis oris activity. Shifting of stress from the first vowel (Fig. 10a) to the second vowel (Fig. 10b) does not disrupt this temporal relationship. Indeed, this closely timed interarticulator relationship has been shown for several other hearing speakers (Tuller & Harris, 1980).

Figure 11a shows one of the tokens, perceived as correct, that most closely resembles those of the hearing speaker. Peak genioglossus activity occurs between the second and third peaks of orbicularis oris activity, but the peak is late relative to the acoustic event. Timing between the articulators differs from the hearing speaker in that genioglossus activity begins after the second orbicularis oris peak occurs, and continues well into the third burst of orbicularis oris activity.

Figure 11b shows a token perceived as a stress error. Genioglossus activity begins quite late relative to orbicularis oris activity, and in fact, it peaks simultaneously with the third oribularis oris peak. This pattern was never seen for the hearing speaker.

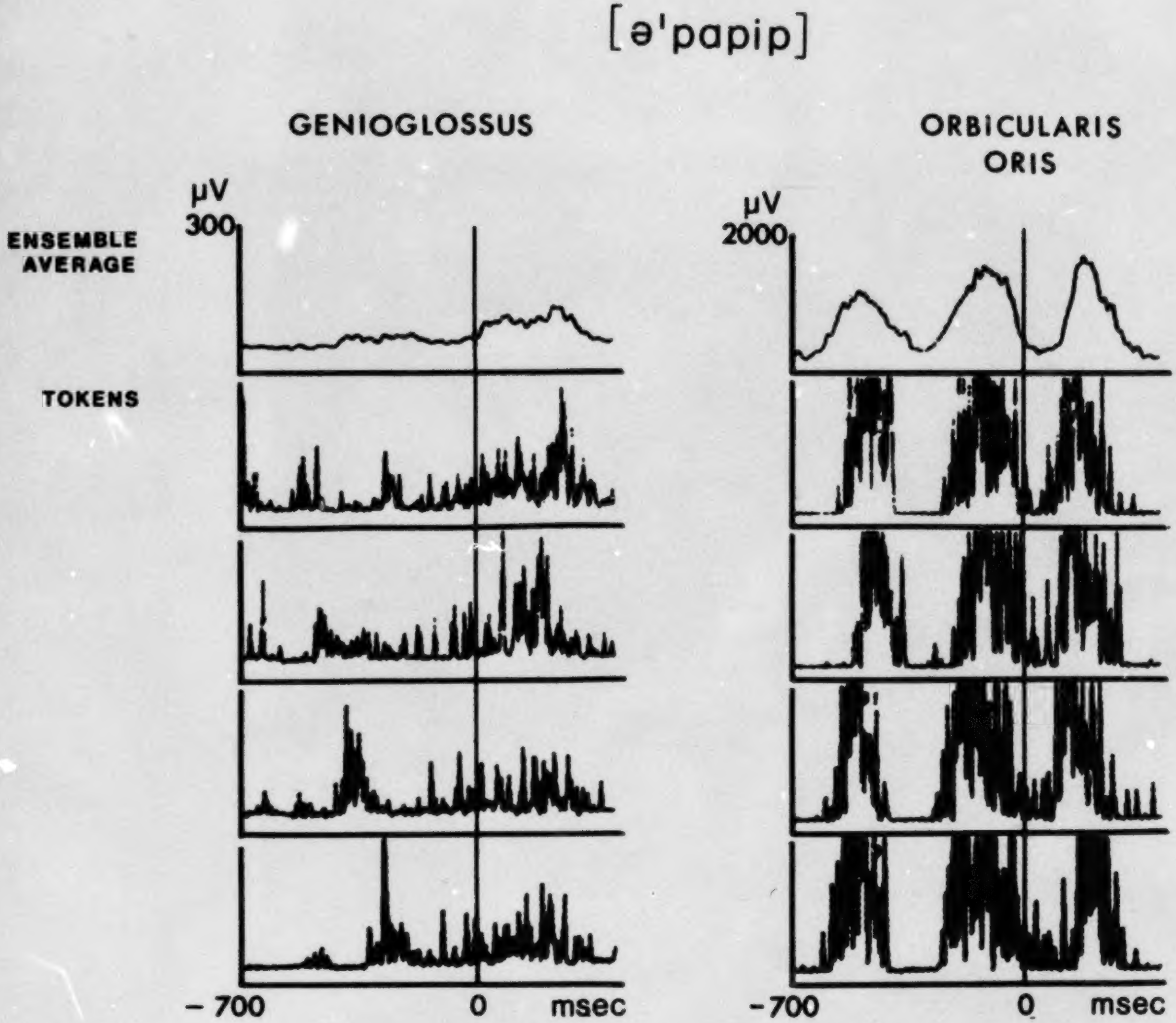


Figure 8. /ə'pɑpip/ as produced by a deaf speaker. Data presented as in Figure 6.

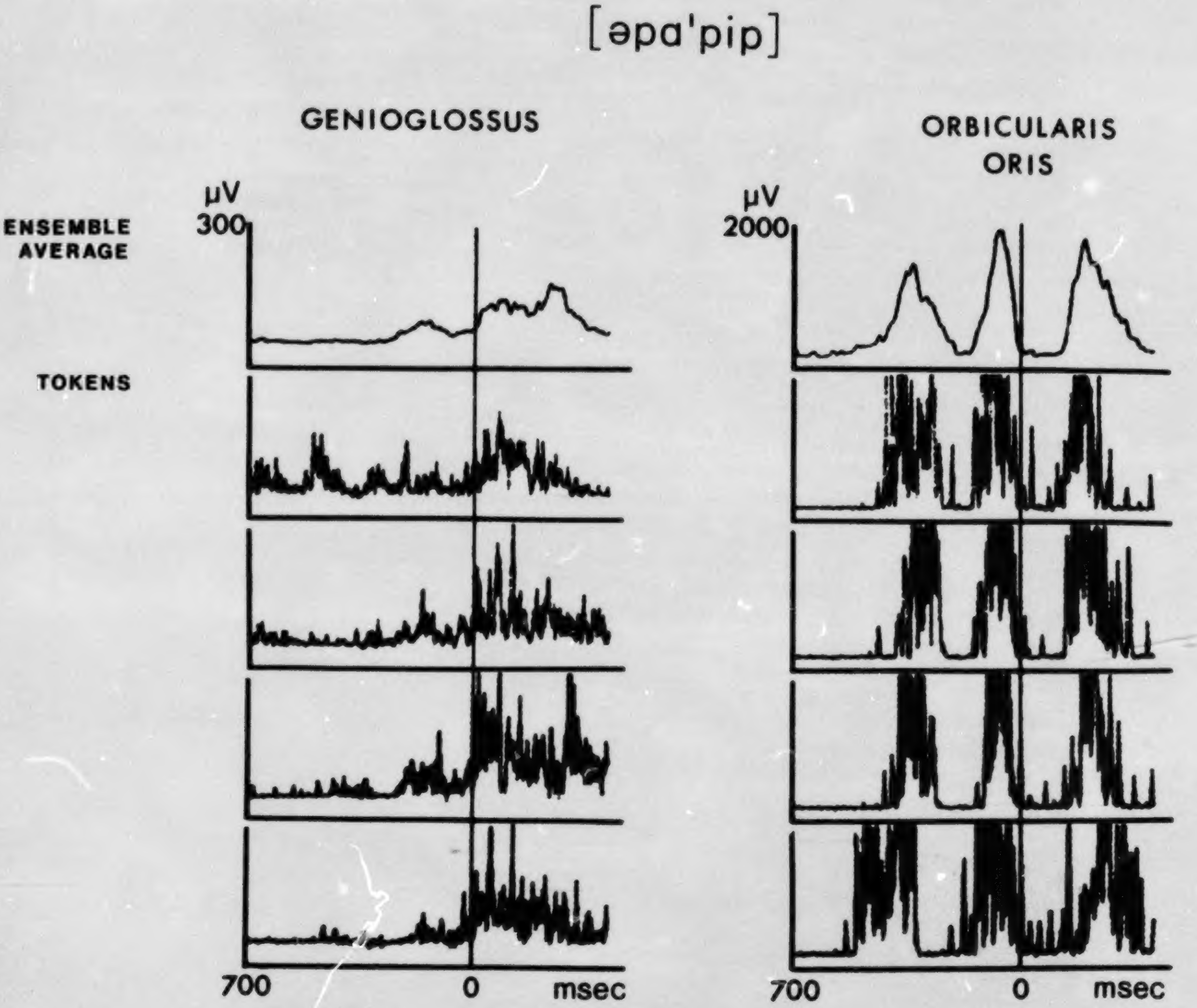


Figure 9. /əpɑ́pip/ as produced by a deaf speaker. Data presented as in Figure 6.

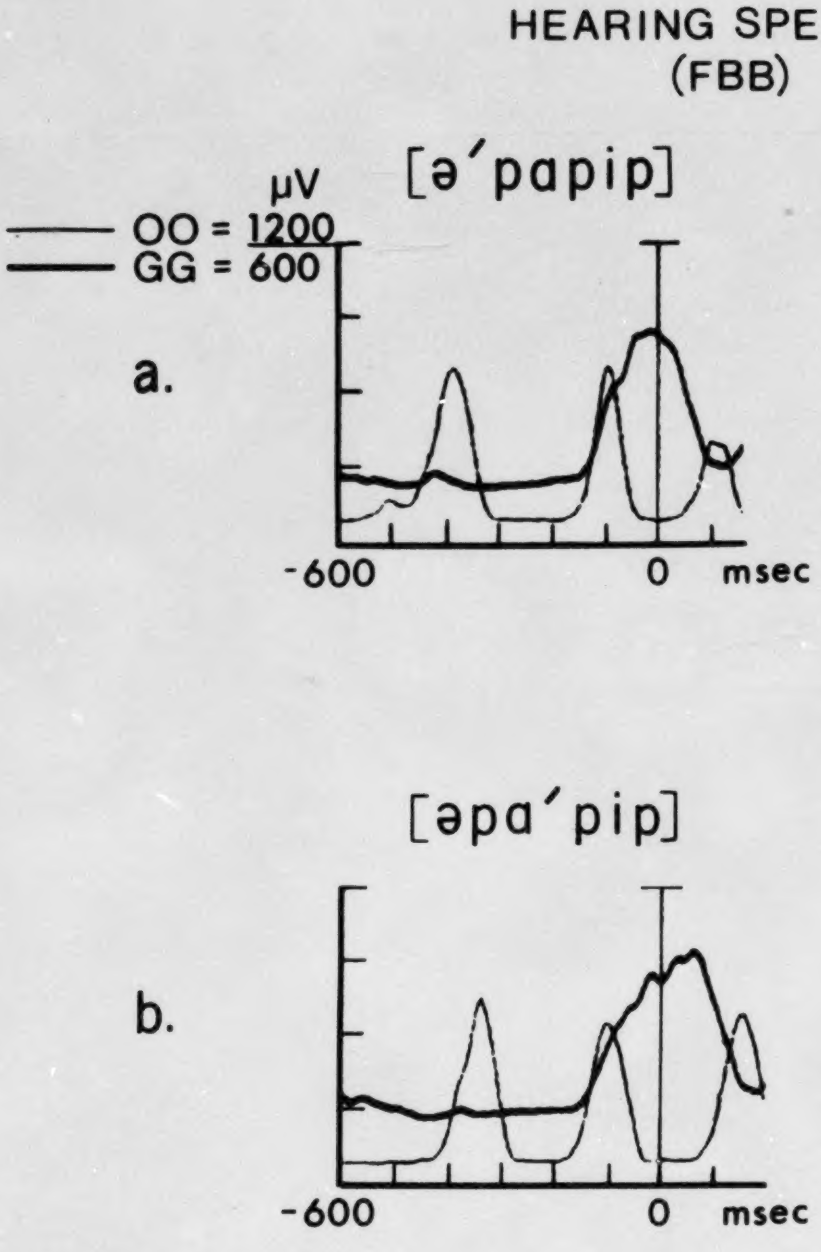


Figure 10. Ensemble average of the EMG potentials for genioglossus and orbicularis oris for the utterance type /əpəpip/ produced by the hearing speaker. The vertical line indicates the acoustic release of the /p/ closure.

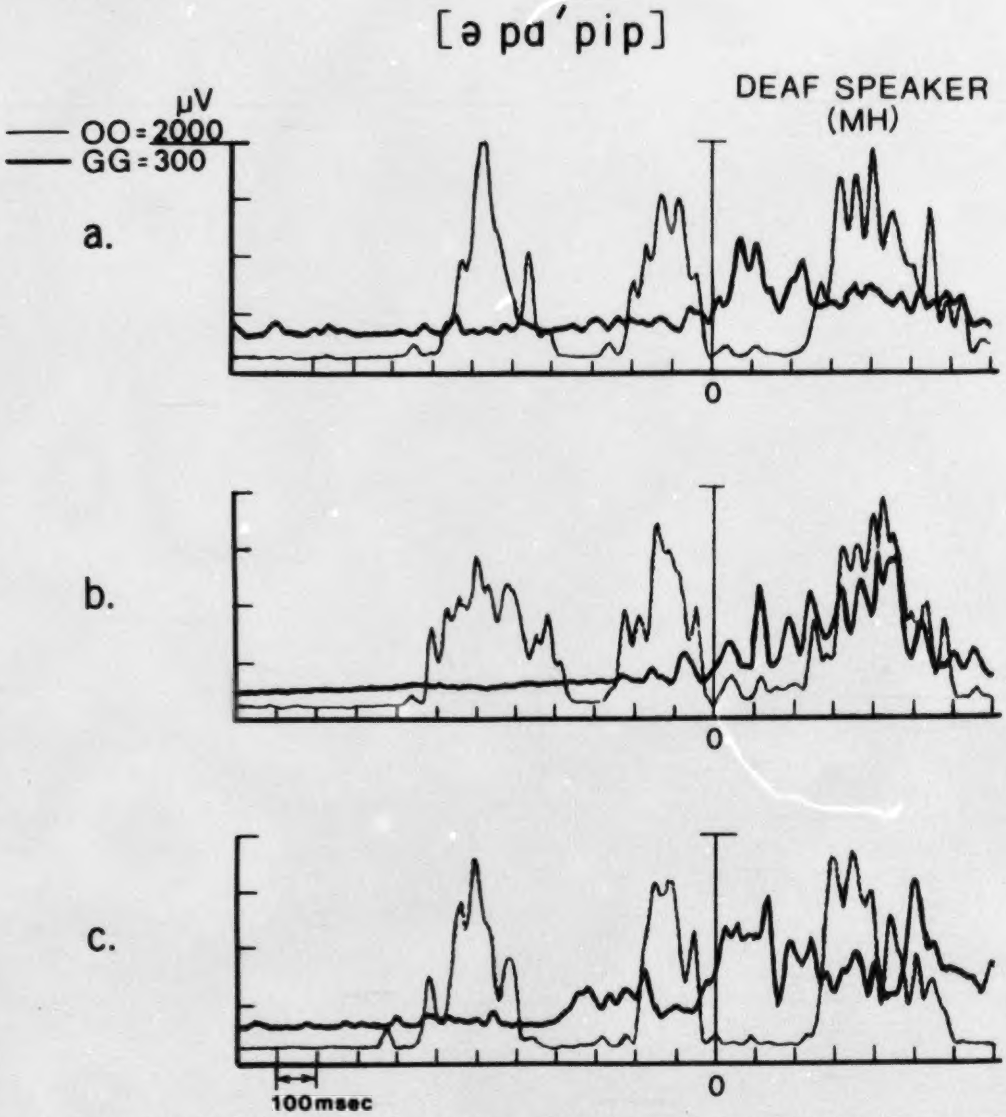


Figure 11. A single selected token of the EMG potential from the genioglossus and orbicularis oris muscles as produced by the deaf speaker. The vertical line indicates the acoustic release of the /p/ closure. In Figure 11a, peak genioglossus activity occurs between the second and third orbicularis oris peaks, but is late relative to the acoustic event. This pattern was most like normal. In Figure 11b and 11c, the single tokens show that genioglossus activity was either too late or too early respectively. (N.B. Single tokens filtered with settings used for the average in Figure 8.)

Figure 11c shows another token perceived as a stress error. Genioglossus activity begins too soon in this case, although a peak occurs between the second and third peaks of orbicularis oris activity. However, the genioglossus activity continues beyond the final burst of orbicularis oris activity.

Figures 12a and 12b show respective examples of: (1) a perceived vowel error, and (2) an instance in which there was inappropriate genioglossus activity for the /ɑ/, but listeners perceived the vowel as correct. These two final examples were quite unusual with respect to the normal. It should be emphasized that while there was substantial token-to-token variation in the deaf speaker, the types of physiological patterns do not differ systematically from "correct" to "incorrect" tokens.

DISCUSSION

While this study obviously does not allow definitive answers to questions about other deaf speakers, it does suggest some further directions for research. First, these results give ample evidence of the instability of deaf production. The speaker does not produce a "wrong" pattern in a stereotyped way; rather, production is variable in all acoustic and physiological measurements we examined. If the results for this speaker are replicated in further work, we cannot assume the deaf speaker simply operates in a reduced or deviant phonological space, whether the distortion of phonology is produced by explicit teaching or some other aspect of the speaker's experience. While the instability has been noted in transcription studies (e.g., Oller & Eilers, in press), it is better documented by studies that go beyond traditional techniques (Fisher, King, Parker, & Wright, in press).

At a segmental level, there is an apparent failure of consistent interarticulator programming. Overall, a tight temporal coupling of activity in articulatory muscles is lacking. For the normal hearing speaker producing a stop consonant-vowel syllable, activity of the tongue muscles for the vowel is well underway when acoustic release for the stop takes place--this may not be so in this deaf speaker. However, the more important difference between deaf and normal subjects is that the relationship between lip and tongue activity varies from token-to-token in the deaf speaker. It is interesting that the variability of the relationship arises from the lingual rather than the labial component--that is, it is the invisible rather than the visible aspect of articulation that varies.

The second hypothesis about deaf speech, described above, is that the tongue is relatively immobile in this group, as inferred from acoustic measures of formant positions, and this contributes to the unintelligibility of the speech (Monsen, 1976). This hypotheses is, in some sense, an extension of the common observation that deaf vowels are neutralized. When we examine our deaf speaker's data, we note that she is capable of contracting an appropriate muscle for /i/, and leaving it relatively inactive for /ɑ/. Thus, the tongue cannot be in the same position for the two vowels. Of course, the present EMG technique cannot be used to ascertain absolute tongue position. The absolute level of EMG activity is not interpretable, since, in addition to the relative strength of muscle contraction, the amplitude of recorded EMG activity reflects the distance of the active electrode from the firing muscle

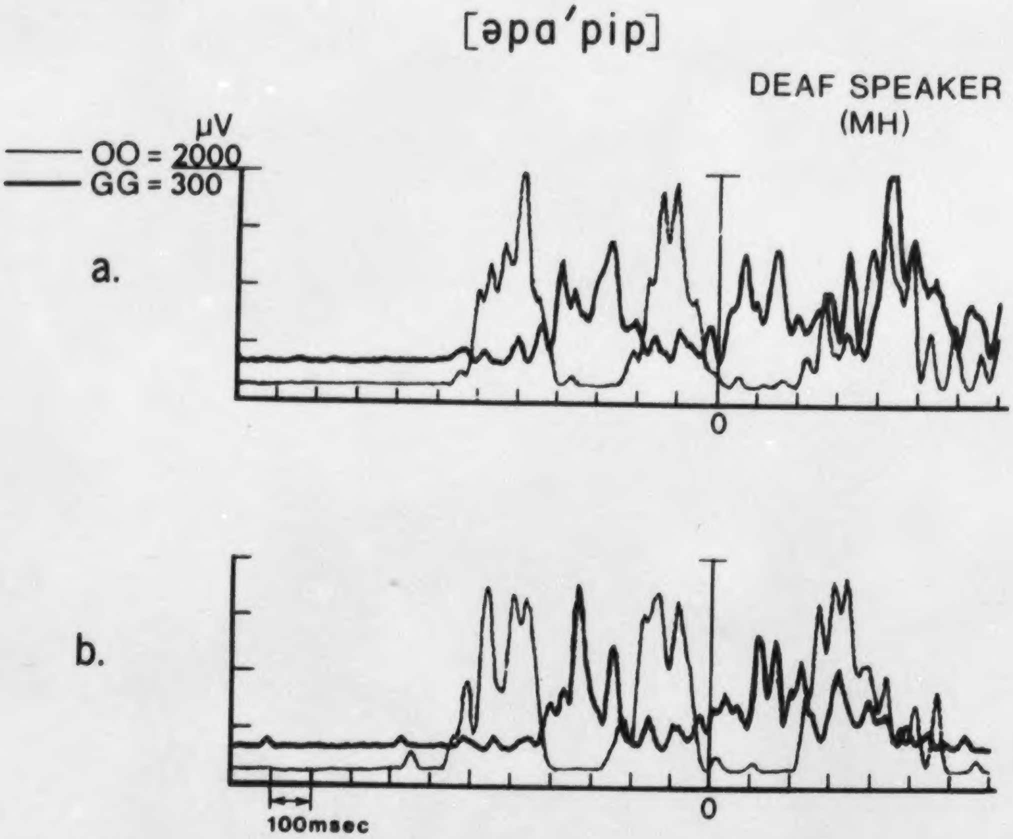


Figure 12. Figure 12a shows an example of a perceived vowel error, with genioglossus activity occurring between the first and second orbicularis oris peaks. This token was perceived as /əpipip/. Figure 12b shows an example of an utterance perceived as correct although genioglossus activity clearly occurs between the first and second orbicularis oris peaks as seen above. Data after Figure 11.

fibers. With respect to the vowel neutralization, we note that her formant values for /i/ and /ɑ/ are more similar to each other than those for the "average" female speaker of Peterson and Barney (1952).

A third hypothesis about deaf speech is that source function control is a substantial source of unintelligibility. The present speaker apparently knew the rules for conveying stress by varying F_0, duration, and intensity, even though she showed the characteristic overall durational lengthening of deaf speech. What is puzzling is that listeners were not able to extract this information from the signal, as shown by the similarity of "correct" and "incorrect" tokens in acoustic measures. We examined the possibility that "incorrect" tokens were those in which conflicting cues were presented, but no such readily apparent pattern emerged. It is possible that the contours of intensity and F_0 were abnormal although the syllable center values were in appropriate ratio.

A question we could not answer within the framework of the present study is what contribution source function irregularities may contribute to segmental unintelligibility. The present experiment suggests an articulatory variable, interarticulator timing, which deserves greater attention. However, it would also be interesting to know how much a deviant and inadequate source in and of itself prevents the listener from interpreting the segmental cues that are received, however inadequate they may be. We intend to pursue this question further, by examining simple nonsense syllables within a wider range of phonetic structures, attempting to use various instrumental techniques to manipulate the source function.

REFERENCE NOTE

1. Levitt, H., Stark, R., McGarr, N. S., Carp, J., Stromberg, H., Gaffney, R. S., Barry, C., Velez, A., Osberger, M. J., Leiter, E., & Freeman, L. Language and Communication Skills of Deaf Children 1973-1976, unpublished report.

REFERENCES

Allen, G. D., Lubker, J. F., & Harrison, E. New paint-on electrodes for surface electromyography. Journal of the Acoustical Society, 1972, 124. (Abstract)

Angelocci, A., Kopp, G., & Holbrook, A. The vowel formants of deaf and normal hearing eleven to fourteen year old boys. Journal of Speech and Hearing Disorders, 1964, 29, 156-170.

Boothroyd, A. Distribution of hearing levels in the student population of Clarke School for the Deaf. SARP Report 3. Northampton: Clarke School for the Deaf, 1970.

Calvert, D. R. Some acoustic characteristics of the speech of profoundly deaf individuals. Unpublished doctoral dissertation, Stanford University, 1961.

Fisher, J., King, A., Parker, A., & Wright, R. Assessment of speech production and speech perception as a basis for speech therapy. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), *Speech of the hearing impaired: Research, training and personnel preparation*. Washington, D.C.: A. G. Bell Association, in press.

Forner, L. L., & Hixon, T. J. Respiratory kinematics in profoundly hearing impaired speakers. *Journal of Speech and Hearing Research*, 1977, *20*, 373-497.

Fry, D. B. Experiments in the perception of stress. *Language and Speech*, 1958, *1*, 126-152.

Fry, D. B. The dependence of stress judgements on vowel formant structure. In *Proceedings of the Fifth International Congress of Phonetic Sciences*. Munster: S. Karger, 1964, 306-311.

Harris, K. S. Vowel duration change and its underlying physiological mechanisms. *Language and Speech*, 1978, *21*, 354-361.

Harris, K. S., & McGarr, N. S. Relationships between speech perception and speech production in normal and hearing-impaired speakers. *Haskins Laboratories Status Report on Speech Research*, 1980, *SR-61*, 23-46.

Haycock, G. S. *The teaching of speech*. Washington, D.C.: Volta Bureau, 1933.

Hirose, H. Electromyography of the articulatory muscles: Current instrumentation and techniques. *Haskins Laboratories Status Report on Speech Research*, 1971, *SR-25/26*, 73-86.

Hudgins, C. V., & Numbers, F. C. An investigation of the intelligibility of the speech of the deaf. *Genetic Psychological Monograph*, 1942, *25*, 289-392.

Huggins, A. W. F. Better spectrograms from children's speech. *Journal of Speech and Hearing Research*, 1980, *23*, 19-27.

Huntington, D. A., Harris, K. S., & Sholes, G. An electromyographic study of consonant articulation in hearing impaired and normal speech. *Journal of Speech and Hearing Research*, 1968, *11*, 147-158.

Johnson, D. Communication characteristics of NTID students. *Journal of the Academy of Rehabilitative Audiology*, 1975, *8*, 17-32.

Kewley-Port, D. Computer processing of EMG signals at Haskins Laboratories. *Haskins Laboratories Status Reports on Speech Research*, 1973, *SR-33*, 173-183.

Levitt, H., Smith, C. R., & Stromberg, H. Acoustic, articulatory, and perceptual characteristics of the speech of deaf children. In G. Fant (Ed.), *Proceedings of the Speech Communication Seminar Stockholm*. New York: John Wiley and Sons, 1974, 129-139.

Lindblom, B. E. F. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 1963, *35*, 1773-1781.

Ling, D. *Speech and the hearing impaired child: Theory and practice*. Washington, D.C.: A. G. Bell Association, 1976.

McGarr, N. S. *The differences between experienced and inexperienced listeners in understanding the speech of the deaf*. Unpublished doctoral dissertation, The City University of New York, 1978.

McGarr, N. S., & Osberger, M. J. Pitch deviancy and intelligibility of deaf children's speech. *Journal of Communication Disorders*, 1978, *11*, 237-247.

Monsen, R. B. Durational aspects of vowel production in the speech of deaf children. *Journal of Speech and Hearing Research*, 1974, *17*, 386-398.

Monsen, R. B. Second formant transitions of selected consonant-vowel combinations in the speech of deaf and normal hearing children. Journal of Speech and Hearing Research, 1976, 19, 279-289.

Oller, D. K., & Eilers, R. E. A pragmatic approach to phonological systems of deaf speakers. In N. Lass (Ed.), Speech and language: Advances in basic research and practice. New York: Academic Press, in press.

Osberger, M. J. & Levitt, H. The effect of timing errors on the intelligibility of deaf children's speech. Journal of the Acoustical Society of America, 1979, 66, 1316-1324.

Peterson, G. E., & Barney, H. L. Control methods used in a study of the identification of vowels. Journal of the Acoustical Society, 1952 32, 693-703.

Raphael, L. J., & Bell-Berti, F. Tongue musculature and the feature of tension in English vowels. Phonetica, 1975, 32, 61-73.

Raphael, L. J., Bell-Berti, F., Collier, R., & Baer, T. Tongue position in rounded and unrounded front vowel pairs. Language and Speech, 1979, 22, 37-48.

Smith, C. R. Residual hearing and speech production of deaf children. Unpublished doctoral dissertation, The City University of New York, 1972.

Tuller, B., & Harris, K. S. Temporal models of interarticulator programming. Journal of the Acoustical Society of America, 1980, 67. (Abstract)

Whitehead, R. Some respiratory and aerodynamic patterns in the speech of the hearing-impaired. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.), Speech of the hearing impaired: Research, training and personnel preparation. Washington D.C.: A. G. Bell Association, in press.

FOOTNOTE

[1] For convenience in the ensuing discussion, we will call the speech characteristic of the group "deaf speech" and for the purposes of the paper, speakers of "deaf speech" will be called deaf. By making this identification, we wish to acknowledge the fact that persons who are severely to profoundly hearing impaired do not necessarily produce this characteristic speech.

ACOUSTIC FACTORS THAT MAY CONTRIBUTE TO CATEGORICAL PERCEPTION*

Janet G. May

Abstract. The perception of the voiced and voiceless velar and pharyngeal fricatives /ɣ, x, ʕ, ħ/ and of /š, s/ in Colloquial Egyptian Arabic was examined to determine if the presence of the first two or three formants in /ɣ, x, ʕ, ħ/ results in continuous perception, in contrast to an expected categorical perception of /š, s/, which lack these formants. Three twelve-step series of VFV nonsense words were synthesized. For the /š/-/s/ series, the center of a band of high-frequency noise was varied in equal steps. For the /x/-/ħ/ and /ɣ/-/ʕ/ series, F1 was varied. Eight native speakers were asked to identify the stimuli and discriminate two-step differences in a 4IAX discrimination task. While the voiced /ɣ/-/ʕ/ series showed continuous or less categorical perception than the /š/-/s/ series, the voiceless /x/-/ħ/ series was perceived somewhat categorically. This suggests that voicing alone, or in combination with acoustic information about the lower formants, may be a necessary condition for continuous perception.

INTRODUCTION

Although the past thirty years have witnessed a revolution in speech research, one of the earliest discoveries made about speech perception still remains somewhat of a mystery: the finding that some speech sounds are perceived in a manner quite different from others. Stop consonants are usually perceived categorically: Subjects can only discriminate as many sounds as they have different labels for (Liberman, Harris, Hoffman, & Griffith, 1957). On the other hand, vowels are perceived more or less continuously: Subjects can discriminate acoustic differences between phonetically equivalent stimuli (Fry, Abramson, Eimas, & Liberman, 1962).

However, categorical perception is not speech-specific (see Strange & Jenkins, 1978). It has been demonstrated for such psychophysical continua as noise-buzz sequences, tone onset times, and visual flicker fusion (Miller,

---

*This paper is based upon a 1979 University of Connecticut doctoral dissertation entitled "The Perception of Egyptian Arabic Fricatives." A shorter version of this paper was presented at the 97th Meeting of the Acoustical Society of America, Boston, Spring 1979.

Acknowledgment. This research and the preparation of this manuscript were supported by NICHD Grant HD-01994 and BSR Grant RR-05596 to the Haskins Laboratories. Many thanks to Ignatius G. Mattingly, Arthur S. Abramson, and Bruno H. Repp for their helpful advice at all stages of work.

[HASKINS LABORATORIES: Status Report on Speech Research SR-63/64 (1980)]

Weir, Pastore, Kelly, & Dooling, 1976; Pisoni, 1977; Pastore, Ahroon, Baffuto, Friedman, Puleo, & Fink, 1977). In addition, the degree of categorical perception can be manipulated by training, experience, task variables, interstimulus relations, and other experimental factors. For example, subjects can be trained to perceive voicing and place features in stop consonants non-categorically (Barclay, 1972; Carney, Widin, & Viemeister, 1977; Samuel, 1977). If vowels are shortened, put in CVC syllables, or degraded by adding noise to them, they show a tendency for categorical perception (Lane, 1965; Stevens, 1968; Sachs, 1969; Fujisaki & Kawashima, 1968). And increasing the interstimulus interval will cause an increase in the degree of categorical perception (Pisoni, 1971).

To account for the perceptual difference between stop consonants and vowels, Fujisaki and Kawashima (1969, 1970, 1971a, 1971b) proposed a model of speech perception in an experimental situation. They suggested that when a subject hears a speech stimulus, he stores two kinds of information about it in short term memory: an echoic memory containing information about the acoustic details of the sound, and a phonetic memory containing a phonetic label. Due to its discrete nature, phonetic memory will endure longer than echoic memory. Furthermore, since stop consonants are short, their echoic memories will decay rapidly, and therefore may not be available to enable a subject to discriminate phonetically equivalent stimuli. Consequently, he or she will have to refer to labels stored in phonetic memory that will allow discrimination of only as many stimuli as the subject has different labels for. Since vowels are much longer in duration, their echoic memories will persist longer than those for stops, and will probably be available when a subject needs them. The information in echoic memory will allow the subject to discriminate acoustic differences between phonetically equivalent stimuli. This would explain why stops are perceived categorically and why vowels are perceived continuously.[1]

There is some reason to believe that this difference in the echoic memories of stop consonants and vowels is due to their differential durations. If, indeed, long duration is a necessary condition for continuous perception, it is certainly not a sufficient condition. The fricatives /š/ and /s/, which can have durations comparable to those of vowels, are perceived categorically (Fujisaki & Kawashima, 1968, 1969; Repp, 1980). In the production of /š/ and /s/, free zeros created by the cavity behind the constrictional source cancel the lower formant frequencies from the spectra of these fricatives. Perhaps the absence of these formants causes categorical perception by somehow making the echoic memory unreliable, and therefore not available to the subject.

Colloquial Egyptian Arabic offers the opportunity to test this hypothesis, since its phonetic inventory contains fricatives produced in both the front and back cavities of the vocal tract. The front cavity fricatives are the familiar /š/ and /s/. The back cavity fricatives are the less familiar voiced and voiceless velars /ɣ/ and /x/, respectively, and the voiced and voiceless pharyngeals /ʕ/ and /ħ/, respectively. In the production of these back cavity fricatives, the constrictional source is close to the glottis, making the cavity behind the source very short. Such a tube produces anti-resonances with frequencies too high to zero out the lower formants. It was hypothesized that the presence of distinctive lower formants would allow continuous perception of these fricatives by making the echoic memory more dependable.

Recordings were made of a native speaker of Colloquial Egyptian Arabic producing the fricatives /š, s, x, γ, ħ, ʕ/ in intervocalic position. These were used as models for creating synthetic counterparts, which were then presented to subjects for identification and discrimination.

Method

Stimuli. Three twelve-step series of VFV stimuli were created on a Glace-Holmes terminal analog synthesizer (Glace, 1968). The first was a series from /š/ to /s/, the second from /x/ to /ħ/, and the third from /γ/ to /ʕ/.

All stimuli in each series contained the same initial and final /ə/, which was 140 msec long and contained appropriate formant frequency transitions to steady-state segments representing the intervocalic fricatives. In its initial steady-state this vowel had an F1 of 658 Hz, an F2 of 1521 Hz, and an F3 of 2329 Hz.

Each fricative segment in the /š/-/s/ series (Figure 1) was 220 msec long and consisted of a band of high-frequency noise, whose center frequency increased from 2974 Hz for /š/ to 4784 Hz for /s/ in steps of about 165 Hz. Sixty msec transitions for F1, F2, and F3 occurred in the vocalic segments starting with the vowel's steady-state values and ending with 440, 1845, and 2652 Hz, respectively, for /š/, and 440, 1764, and 2652 Hz, respectively, for /s/. Thus, only the F2 transition varied across the series.

Each fricative segment in the /x/-/ħ/ series (Figure 2) was 200 msec long and consisted of the first two noise-excited vocalic formants and a band of high-frequency noise. For all stimuli the second formant was 1886 Hz, and the center of the band of noise was 3961 Hz. The first formant increased from 368 Hz for /x/ to 900 Hz for /ħ/ in steps of about 50 Hz. The amplitude of the high-frequency noise decreased from -24 dB (with respect to the amplitude of the vowel's first formant) for /x/ to -39 dB for /ħ/. Thirty msec transitions for F1, F2, and F3 occurred in the vocalic segments starting with the vowel's steady-state values and ending with 465, 1764, and 2248 Hz, respectively, for /x/, and 827, 1764, and 2248 Hz, respectively, for /ħ/. Thus, only the F1 transition varied across this series.

Each fricative segment in the /γ/-/ʕ/ series (Figure 3) was 110 msec long, and consisted of three vocalic formants and a band of high-frequency noise. For all these segments the second formant was 1521 Hz, the third formant, 2248 Hz, and the center of the band of noise, 3961 Hz. The first formant increased from 368 Hz for /γ/ to 900 Hz for /ʕ/ in steps of about 50 Hz. The amplitude of the high-frequency noise was decreased from -13 dB for /γ/ to -39 dB for /ʕ/. The vocalic formants and the band of noise were synthesized using a mixture of periodic and aperiodic excitation. The ratio of periodic to aperiodic excitation increased with each step along the series. This was achieved by interspersing an increasing number of 10 msec intervals of periodic excitation among a decreasing number of 10 msec intervals of aperiodic excitation, until the last stimulus in the series contained only periodic excitation during this segment. Fifty msec transitions for F1, F2, and F3 occurred in the vocalic segments starting with the vowel's steady-state

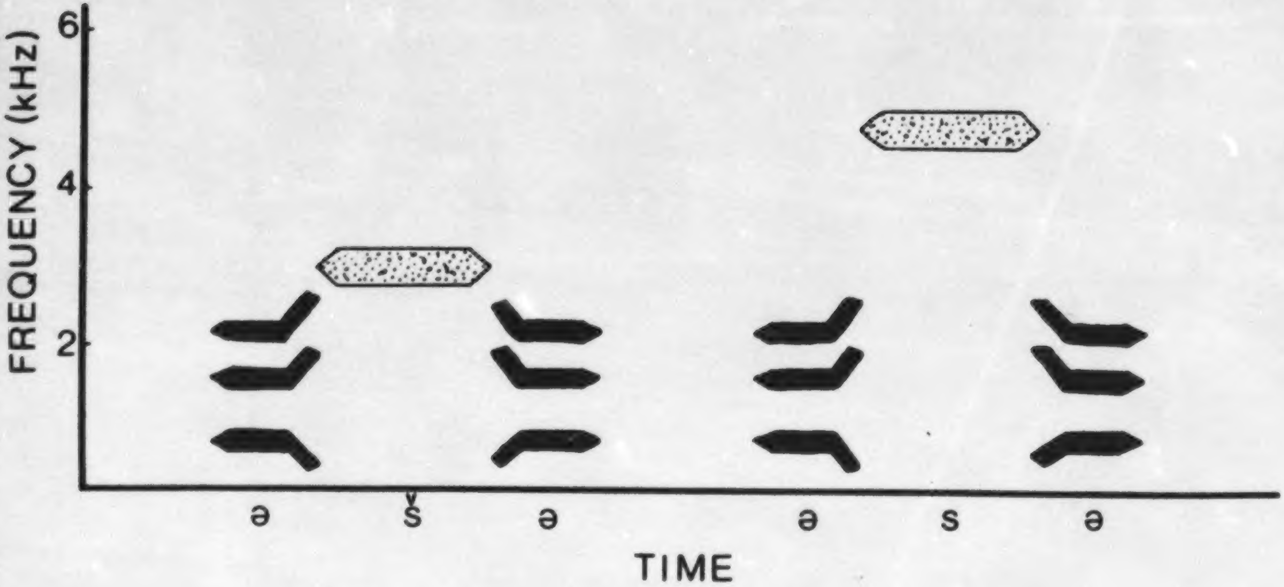

6
4
2
FREQUENCY (kHz)
ə
š
ə
ə
s
ə
TIME

Figure 1. Schematic spectrograms of stimulus #1 (on left) and of stimulus #12 (on right) in the synthetic /š/-/s/ series.

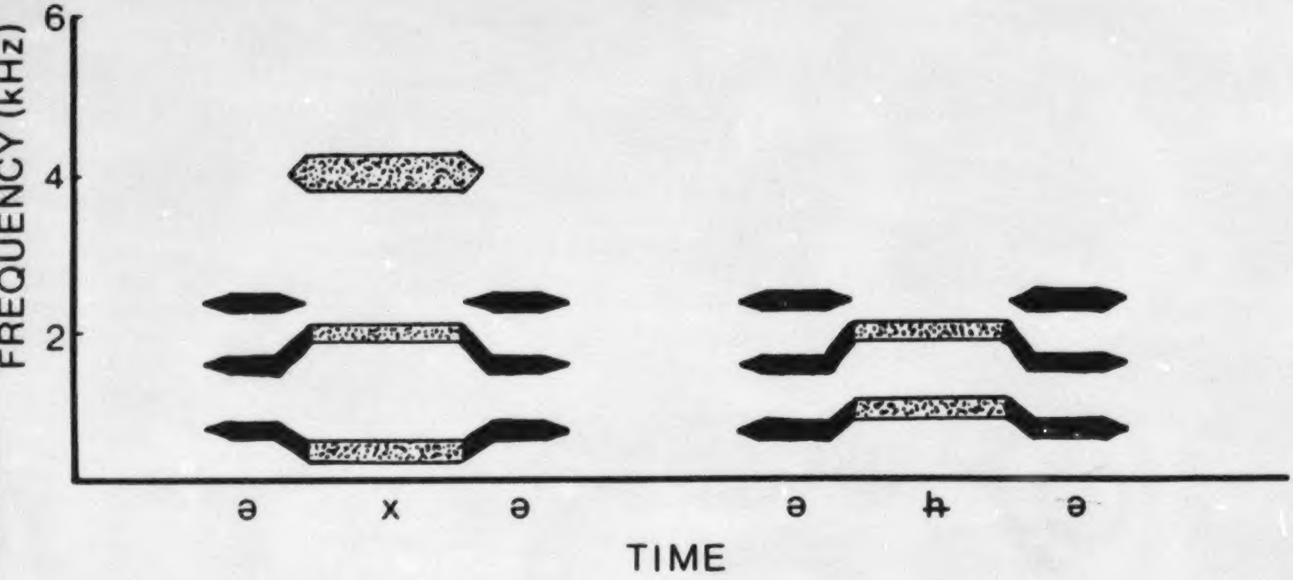

6
4
2
FREQUENCY (kHz)
ə
x
ə
ə
ħ
ə
TIME

Figure 2. Schematic spectrograms of stimulus #1 (on left) and of stimulus #12 (on right) in the synthetic /x/-/ħ/ series.

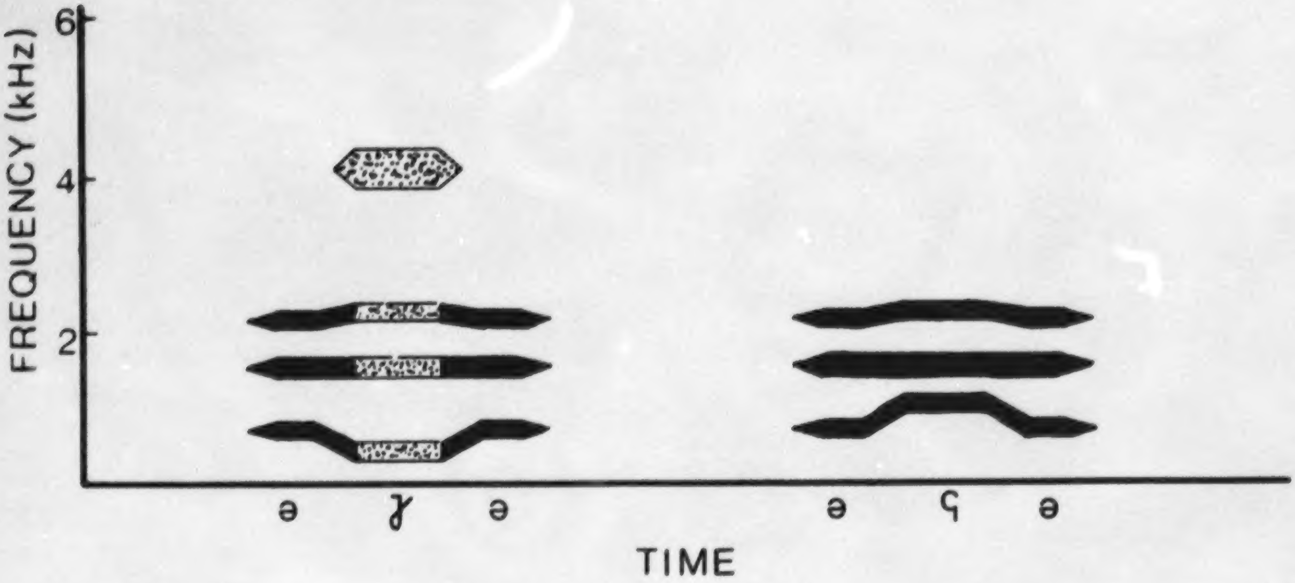

6
4
2
FREQUENCY (kHz)
ə
ɣ
ə
ə
ʕ
ə
TIME

Figure 3. Schematic spectrograms of stimulus #1 (on left) and of stimulus #12 (on right) in the synthetic /ɣ/-/ʕ/ series.

values and ending with 416, 1521, and 2248 Hz, respectively, for /ɣ/, and 852, 1521, and 2248 Hz, respectively, for /ʕ/. Thus, only the F1 transition varied across this series.

Experimental tests. One identification test and three 4IAX discrimination tests were prepared for each series of stimuli. In the identification test, subjects were asked to identify each of the 12 stimuli in a series 16 times. In each of the three discrimination tests subjects were asked to discriminate each two-step difference 8 times, totaling 24 trials across the three tests. The odd stimulus occurred in each position of the 4IAX pairs an equal number of times. A subject responded by writing "1" or "2" to indicate whether the first or second pair of stimuli contained different sounds.

Subjects. Eight phonetically naive adult native speakers of Egyptian Arabic (not including the original native informant), all from Cairo or nearby, were used as paid subjects in these experiments. One subject showed somewhat erratic behavior on the /š/-/s/ identification test, although her discrimination curves for this series showed a peak where one would expect a phoneme boundary. Since discrimination performance predicted from these identification data would be rather irregular, it would be difficult to compare it to the obtained discrimination. In addition, results from most other tests indicate that she was generally an inattentive subject. Consequently, this subject was eliminated from the study.

Procedure. Each subject took twelve tests: one identification test and three discrimination tests for each of the three continua. The subjects were first given all four tests for the /ɣ/-/ʕ/ series, then all tests for the /š/-/s/ series, and finally all tests for the /x/-/ħ/ series. The subjects were divided into two groups of four. Within each group of four tests for a given series, one group of subjects always heard the identification test first, while the other group heard the discrimination tests first. Two tests were administered per experimental session: either one identification test and one discrimination test, or two discrimination tests. Each test took approximately fifteen minutes. The subjects had a brief rest period between the two tests. Their responses for the /š/-/s/ series were very inconsistent. Presumably, this was caused by "clipping" of the signal due to a rather high playback level. Therefore, after all other tests had been administered, the /š/-/s/ identification and discrimination tests were presented to subjects with a reduced playback level for a second time. The results of this second presentation are reported here.

RESULTS

Identification. Individual responses were sufficiently alike to warrant pooling of the data. Pooled identification functions are shown in the top halves of Figures 4-6. Each point represents 112 judgments, 16 per subject. The functions for each of the three series demonstrate that subjects consistently divided each into two discrete categories: /š/ and /s/, /x/ and /ħ/, or /ɣ/ and /ʕ/.

Discrimination. Comparison of the group that took all identification tests first with the group that took all discrimination tests first showed that there was no statistically significant difference (Student's t-test) between the two groups in the discrimination performance for each of the three continua. Therefore, responses from both groups were pooled. In addition, subjects did not exhibit a bias for responding "1" or "2" on any of the discrimination tests (Student's t-test).

Ideal categorical perception is characterized by a subject's ability to discriminate only as many sounds as he can identify, as predicted by Formula 1 (see Pollack & Pisoni, 1971 for derivation):

$$P(C) = \frac{(a-a')^2 + (b-b')^2 + 2}{4} \tag{1}$$

where $P(C)$ represents the probability of correctly discriminating A and B, $a = P(a|A)$ (the probability of labeling stimulus A as phoneme *a*), $a' = P(a|B)$ (the probability of labeling stimulus B as phoneme *a*), $b = P(b|A)$, and $b' = P(b|B)$.[2]

These predictions are represented in the bottom halves of each of the Figures 4-6 by the open circles. Obtained discrimination scores are denoted by the closed circles, each of which represents 168 judgments on the composite function, 24 per subject. The stimulus pair labeled "1" refers to a pair composed of stimuli 1 and 3, etc.

The identification function in Figure 4 shows that the phoneme boundary for the /š/-/s/ series is located between stimuli 6 and 7. Predicted discrimination shows that, if categorical perception obtains, subjects should not be able to discriminate stimulus pairs 1-4, all of whose members are within the /š/ category, and stimulus pairs 7-10, all of whose members are within the /s/ category (50% = chance). Discrimination performance should increase to about 65% for stimulus pairs 5 and 6 whose members are near the phoneme boundary. Obtained discrimination scores are higher than predicted, $F(1,6)=16.1$, $p < .01$, but show a correlation with predicted discrimination. Note that discrimination performance is greatest for stimulus pairs 5 and 6, as predicted.

The identification function in Figure 5 shows that the phoneme boundary for the /x/-/ħ/ series lies close to stimulus 6. Predicted discrimination shows that, if categorical perception obtains, subjects should not be able to discriminate stimulus pairs 1-3, all of whose members lie within the /x/ category, and stimulus pairs 7-10, all of whose members lie within the /ħ/ category. Discrimination performance should increase to about 72% for stimulus pair 5, whose members, namely 5 and 7, straddle the phoneme boundary. Obtained discrimination, though somewhat higher than predicted, $F(1,6)=22.6$, $p < .005$, shows a correlation with predicted discrimination. Discrimination performance increased from 50-60% for stimulus pairs 1 and 2 to 79% for stimulus pair 4, and then decreased to around 60% for stimulus pairs 7-10. Notice that although performance peaks for stimulus pair 5 in the predicted discrimination, it peaks for stimulus pair 4 in the obtained discrimination. However, the members of both these pairs straddle the phoneme boundary, which is located slightly to the left of stimulus 6.

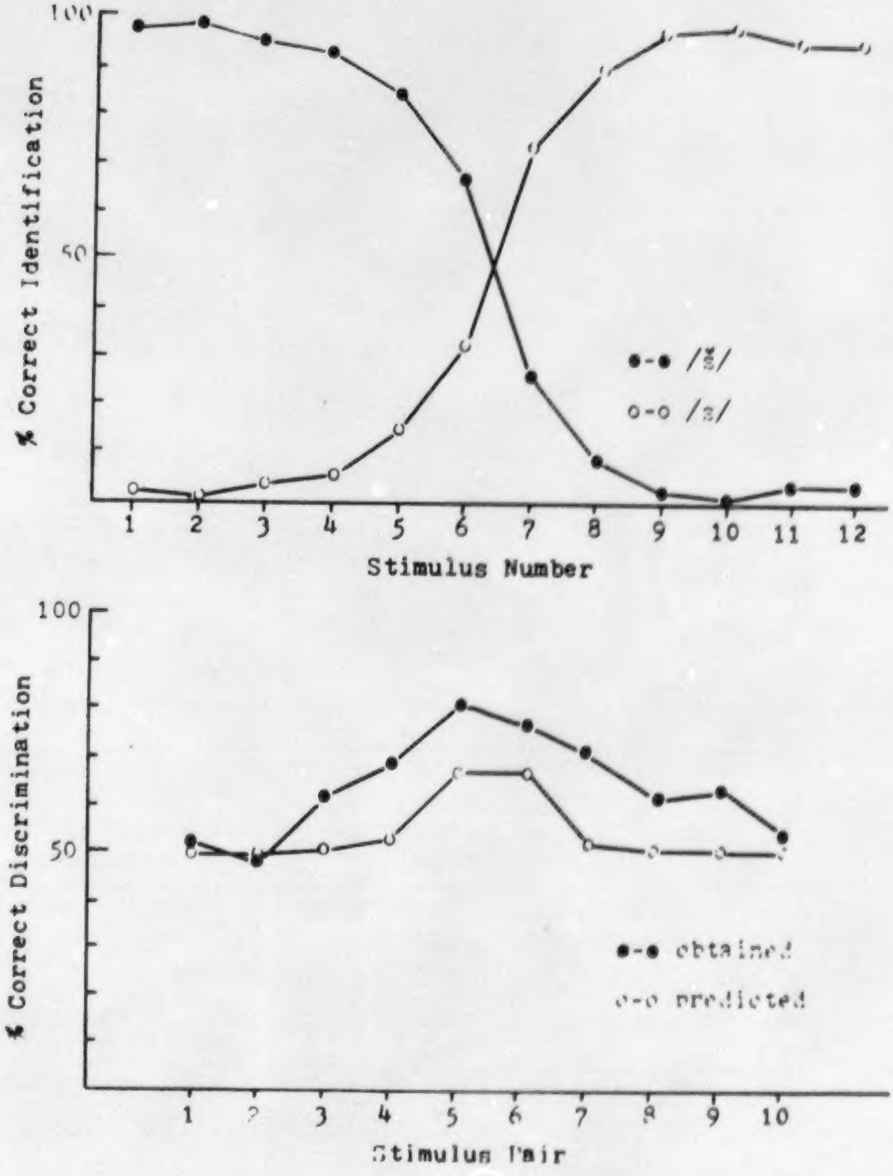


Figure 4. Identification function (top) and predicted and obtained discrimination functions (bottom) for seven subjects for the /š/-/s/ series.

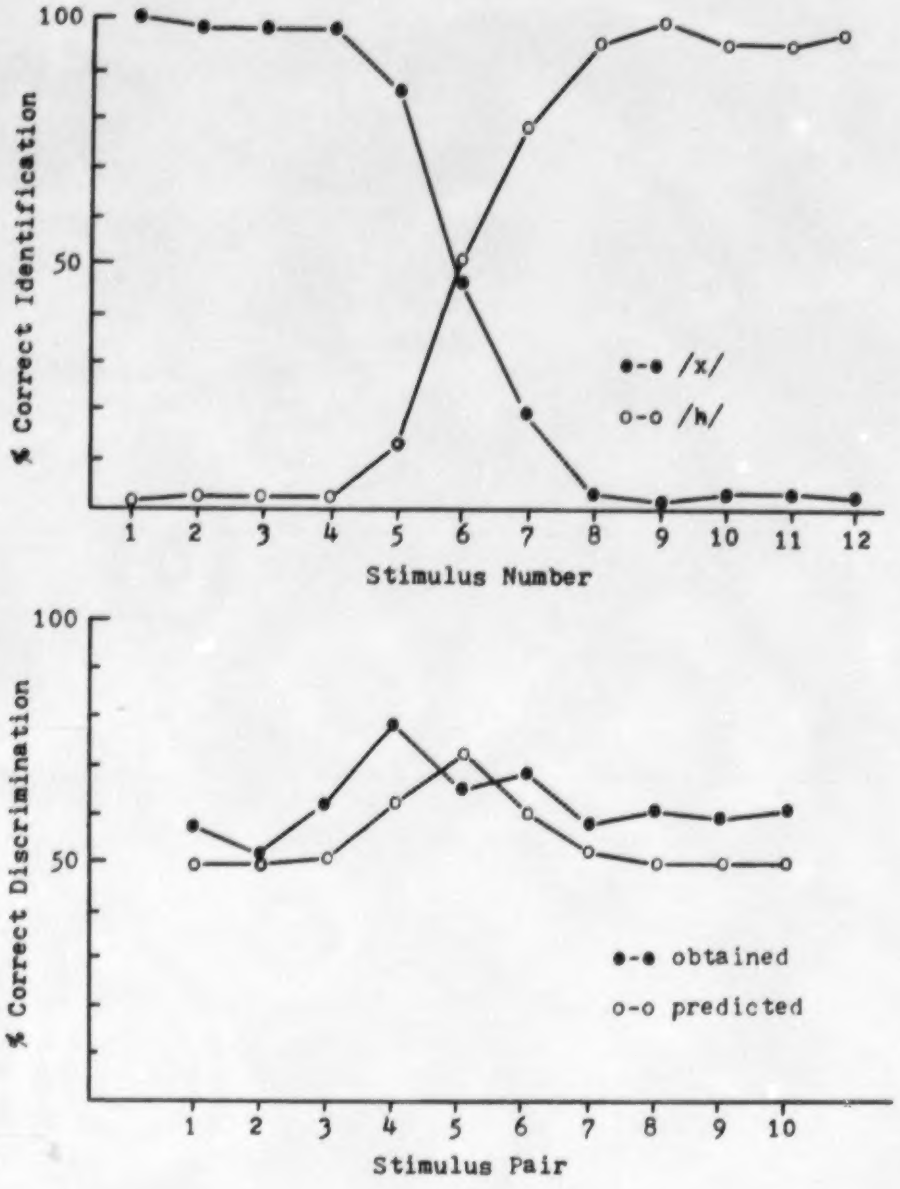


Figure 5. Identification function (top) and predicted and obtained discrimination functions (bottom) for seven subjects for the /x/-/ħ/ series.

The identification function in Figure 6 shows that the phoneme boundary for the /γ/-/ʕ/ series is located between stimuli 7 and 8. Predicted discrimination shows that for categorical perception, subjects should be able to discriminate stimulus pairs 1-5, all of whose members lie within the /γ/ category, and stimulus pairs 8-10, all of whose members lie within the /ʕ/ category, only about 50% of the time. Discrimination performance should increase to about 68% for stimulus pair 6, whose members, namely 6 and 8, straddle the phoneme boundary. Obtained discrimination was significantly greater than predicted discrimination, F(1,6)=142.4, p < .001. Performance increases from about 50% for stimulus pair 2 to about 81% for stimulus pair 4. Performance remains about 70% for stimulus pairs 4-10, and peaks to about 95% for stimulus pair 7, whose members, namely 7 and 9, straddle the phoneme boundary.

These data demonstrate that subjects tend to perceive the voiceless synthetic stimuli in the /š/-/s/ and /x/-/ħ/ series categorically, while they perceive the voiced synthetic stimuli in the /γ/-/ʕ/ series less categorically, or more continuously. An analysis of variance shows this difference to be statistically significant, F(2,12)=12.2, p < .005.

DISCUSSION

The hypothesis examined here is that categorical perception of /š/ and /s/ may, in part, be caused or promoted by a lack of information about the lower formant frequencies in the acoustic signal. It was hypothesized that stimuli in the /š/-/s/ series, which lack these formants, would be perceived categorically, and that stimuli in the /x/-/ħ/ and /γ/-/ʕ/ series, which contain these formants, would be perceived continuously. However, the data from the present experiment show that while subjects indeed perceive the voiced fricatives in the /γ/-/ʕ/ series continuously, and the voiceless fricatives in the /š/-/s/ series categorically, they tend to perceive the voiceless /x/-/ħ/ series categorically. Since all stimuli are of relatively long duration, it cannot be short duration of acoustic cues that is causing categorical perception in this instance. Although these sounds contain information about the acoustic details of the lower formant frequencies, for some reason the echoic stores seem to be unreliable. As a result, subjects cannot use information stored in them to discriminate stimuli, resulting in categorical perception. It is possible that in addition to long duration, noncategorical perception not only requires information about the lower formant frequencies, but also that the stimuli be voiced. In fact, the present data could be explained on the basis of voicing alone: The voiceless fricatives /š, s, x, ħ/ were perceived categorically, and the voiced fricatives /γ, ʕ/ were perceived continuously (just as vowels).

It is interesting to note that results from experiments involving tests of immediate ordered recall of auditorily presented fricatives support this conclusion. In these experiments the voiced fricatives /ž, z, v/, which were presented in isolation and in a CV context, exhibited the recency and suffix effects that had been found earlier for vowels, but not for stop consonants (Crowder, 1973). It is assumed that subjects show significant improvement for recall of the last members of the vowel and voiced fricative series because their echoic memories are more dependable. If this is true, then we would

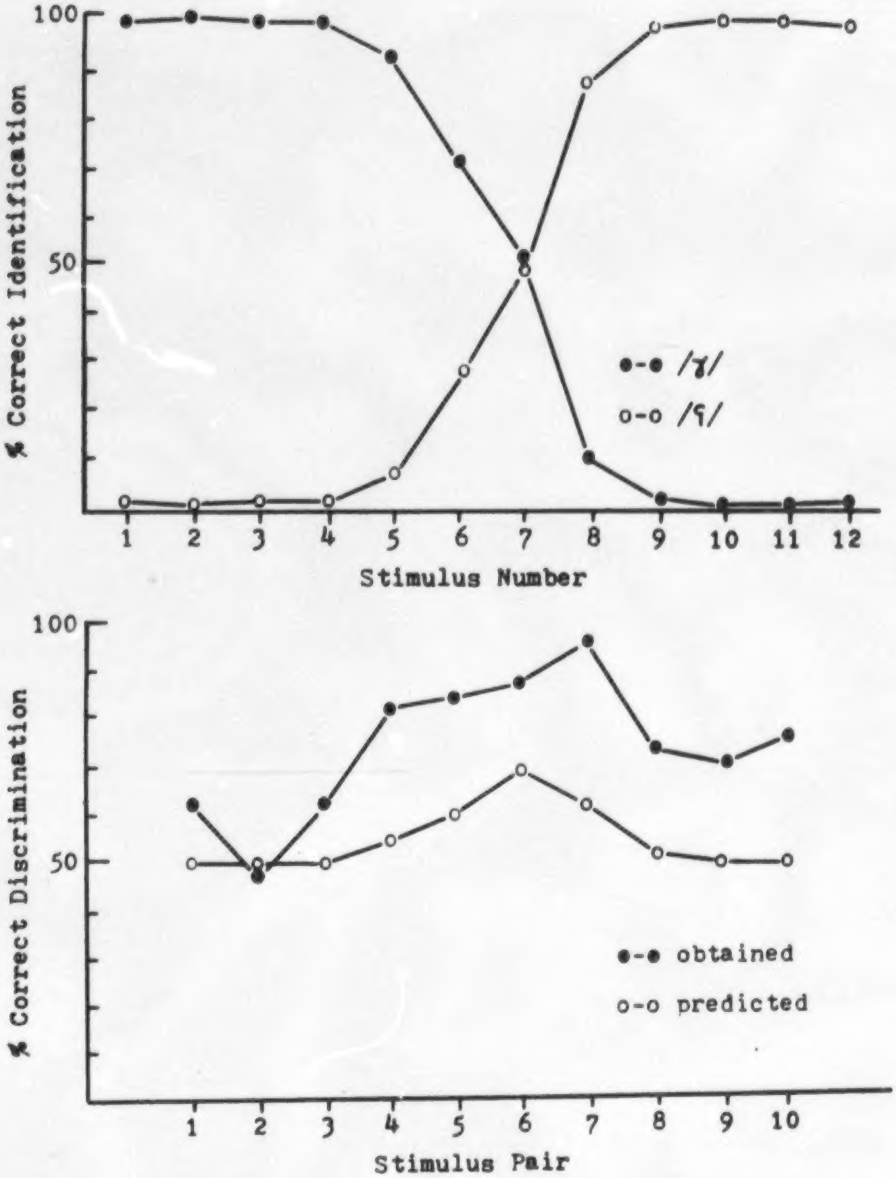


Figure 6. Identification function (top) and predicted and obtained discrimination functions (bottom) for seven subjects for the /ɣ/-/ʕ/ series.

expect subjects to perceive these same stimuli continuously in a discrimination task, because they should be able to refer to echoic memory to help them discriminate stimuli on the basis of differences in the acoustic details of the stimuli.

In conclusion, the results of the experiments in the present study suggest that in addition to cues of long duration, the presence of voicing may be a necessary condition for continuous perception. Since it was found that the voiced fricatives /γ, ʕ/, which contain information about the lower formants, were perceived continuously, but that the voiceless fricatives /x, ħ/, which also contain this information, were perceived categorically, it is unclear whether information about the lower formants contributes to continuous perception, as originally hypothesized. It is hoped that future research involving the perception of /ž, z/ and whispered vowels will shed some light on this matter.

REFERENCES

Barclay, J. R. Noncategorical perception of a voiced stop: A replication. *Perception & Psychophysics*, 1972, *11*, 269-273.

Carney, A. E., Widin, G. P., & Viemeister, N. F. Non-categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, *62*, 961-970.

Crowder, R. G. The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 587-596.

Crowder, R. G. Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology*, 1973, *98*, 14-24.

Crowder, R. G., & Morton, J. Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, *5*, 365-373.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. The identification and discrimination of synthetic vowels. *Language and Speech*, 1962, *5*, 171-189.

Fujisaki, H., & Kawashima, T. The influence of various factors on the identification and discrimination of synthetic speech sounds. *Reports of the Sixth International Congress on Acoustics*, Tokyo, 1968.

Fujisaki, H., & Kawashima, T. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute*, University of Tokyo, 1969, *28*, 67-73.

Fujisaki, H., & Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute*, University of Tokyo, 1970, *29*, 207-214.

Fujisaki, H., & Kawashima, T. A model of the mechanisms for speech perception: Quantitative analysis of categorical effects in discrimination. *Annual Report of the Engineering Research Institute*, University of Tokyo, 1971, *30*, 59-68. (a)

Fujisaki, H., & Kawashima, T. A quantitative model for the mechanisms of discriminating speech sounds. *Reports of the Seventh International Congress on Acoustics*, Budapest, 1971. (b)

Glace, D. A. Parallel resonance synthesizer for speech research. *Journal of the Acoustical Society of America*, 1968, *44*, 391. (Abstract)

Lane, H. L. The motor theory of speech perception: A critical review. *Psychological Review*, 1965, *72*, 275-309.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, *54*, 358-368.

Miller, J. D., Weir, C. C., Pastore, R. W., Kelly, W. J., & Dooling, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, *60*, 410-417.

Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., Puleo, J. S., & Fink, E. A. Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 686-696.

Pisoni, D. B. *On the nature of categorical perception of speech sounds*. Unpublished doctoral dissertation, University of Michigan, 1971.

Pisoni, D. B. Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops. *Journal of the Acoustical Society of America*, 1977, *61*, 1352-1361.

Pollack, I., & Pisoni, D. B. On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, 1971, *24*, 299-300.

Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction: IV. Two strategies in fricative discrimination. *Haskins Laboratories Status Report on Speech Research*, 1980, *SR-63/64*, this issue.

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 129-145.

Sachs, R. M. Vowel identification and discrimination in isolation versus word context. *Quarterly Progress Report* (Research Laboratory of Electronics, Massachusetts Institutes of Technology, Cambridge, Massachusetts), 1969, 220-229.

Samuel, A. G. The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics*, 1977, *22*, 321-330.

Stevens, K. N. On the relations between speech movements and speech perception. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 1968, *213*, 102-106.

Strange, W., & Jenkins, J. J. The role of experience in the perception of speech. In R. D. Walk & H. L. Pick, Jr. (Eds.), *Perception and experience*. New York: Plenum Publishing Corporation, 1978.

FOOTNOTES

[1] There is corroborative evidence for the existence of echoic memory from tests of immediate ordered recall of auditorily presented consonants and vowels. It is assumed that a subject must hold acoustic information about the stimuli in a sensory or prelinguistic form for at least a few seconds until it can be analyzed. This store was termed Precategorical Acoustic Storage (PAS) by Crowder and Morton (1969), and is equivalent to echoic memory. Crowder (1971) found that when subjects are asked to recall a series of vowels, they show a significant improvement on the last few members of the series. This

recency effect was attributed to the existence of PAS for the most recently received vowels, which acts to improve their recall. Since PAS lasts only a few seconds, the PAS for the earlier members of the series will have decayed by the time the subjects are required to recall the series. In addition, when a verbal suffix, which subjects are told to ignore, is added to the end of the series, it seems to interfere with the PAS of vowels and the recency effect is lost. This suffix effect was attributed to interference of the suffix with the PAS of the most recent vowels. It is very interesting to note that neither the recency effect nor the suffix effect was found for the voiced stop consonants. Since stops are relatively short in duration, their PAS may not endure as long as that for vowels. Therefore, the PAS of stop consonants will not be available and so cannot help to improve recall of the last items in the consonant series. Furthermore, a suffix will have nothing to interfere with.

[2] It has been suggested that categorical perception is characterized not only by predictability, but also by absoluteness--the ability to remain unaffected by surrounding context. Therefore, a more accurate measure of degree of continuous perception would involve comparing obtained discrimination with discrimination predicted from an identification test that used the same context (Repp, Healy, & Crowder, 1979). This procedure was brought to my attention too late to be used in these experiments, but will be used in the future.

II. PUBLICATIONS

III. APPENDIX

PUBLICATIONS

Alfonso, P. J., & Daniloff, R. G. Allophonic backward masking of stop consonants. *Phonetica*, in press.

Bell-Berti, F., & Harris, K. S. A temporal model of speech production. *Phonetica*, in press.

Borden, G. J. Use of feedback in established and developing speech. In *Speech and language: Advances in basic research and practice* (Vol. 3). New York: Academic Press, 1980, 223-242.

Goodman, D., & Kelso, J. A. S. Are movements prepared in parts? Not under compatible (naturalized) conditions. *Journal of Experimental Psychology: General*, 1980, *109*, 475-495.

Harris, K. S. Electromyography as a technique for laryngeal investigation. In C. Ludlow (Ed.), *Conference on the assessment of vocal pathology*. ASHA Reports, in press.

Harris, K. S., & McGarr, N. S. Relationships between speech perception and speech production in normal-hearing and hearing-impaired subjects. In J. Subtelny (Ed.), *Speech assessment and speech improvement for the hearing impaired*. Washington, D.C.: A. G. Bell, 1980.

Katz, L., & Feldman, L. B. Linguistic coding in word recognition: Comparisons between a deep and shallow orthography. In C. Perfetti & A. Lesgold (Eds.), *Interactive processes in reading*. Hillsdale, N.J.: LEA, 1981.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, in press.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 1980, *28*, 213-228.

McGarr, N., & Harris, K. S. Articulatory control in a deaf speaker. In H. Levitt (Ed.), *Conference on the speech of the hearing impaired*. Washington, D.C.: A. G. Bell, in press.

Price, P. J. Sonority and syllabicity: Acoustic cues to perception. *Phonetica*, 1980, *37*, No. 6.

Repp, B. H. Accessing phonetic information during perceptual integration of temporally distributed cues. *Journal of Phonetics*, 1980, *8*, 185-194.

Repp, B. H., & Mann, V. A. Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, in press.

APPENDIX

DTIC (Defense Technical Information Center) and ERIC (Educational Resources Information Center) numbers:

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED-044-679
SR-23	July - September 1970	AD 723586	ED-052-654
SR-24	October - December 1970	AD 727616	ED-052-653
SR-25/26	January - June 1971	AD 730013	ED-056-560
SR-27	July - September 1971	AD 749339	ED-071-533
SR-28	October - December 1971	AD 742140	ED-061-837
SR-29/30	January - June 1972	AD 750001	ED-071-484
SR-31/32	July - December 1972	AD 757954	ED-077-285
SR-33	January - March 1973	AD 762373	ED-081-263
SR-34	April - June 1973	AD 766178	ED-081-295
SR-35/36	July - December 1973	AD 774799	ED-094-444
SR-37/38	January - June 1974	AD 783548	ED-094-445
SR-39/40	July - December 1974	AD A007342	ED-102-633
SR-41	January - March 1975	AD A013325	ED-109-722
SR-42/43	April - September 1975	AD A018369	ED-117-770
SR-44	October - December 1975	AD A023059	ED-119-273
SR-45/46	January - June 1976	AD A026196	ED-123-678
SR-47	July - September 1976	AD A031789	ED-128-870
SR-48	October - December 1976	AD A036735	ED-135-028
SR-49	January - March 1977	AD A041460	ED-141-864
SR-50	April - June 1977	AD A044820	ED-144-138
SR-51/52	July - December 1977	AD A049215	ED-147-892
SR-53	January - March 1978	AD A055853	ED-155-760
SR-54	April - June 1978	AD A067070	ED-161-096
SR-55/56	July - December 1978	AD A065575	ED-166-757
SR-57	January - March 1979	AD A083179	ED-170-823
SR-58	April - June 1979	AD A077663	ED-178-967
SR-59/60	July - December 1979	AD A082034	ED181-525
SR-61	January - March 1980	AD A085320	ED-185-636
SR-62	April - June 1980	**	**

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

**DDC and/or ERIC order numbers not yet assigned.